



## FACULTAD DE FARMACIA

Grado en Farmacia

# BÚSQUEDA DE GENES CODIFICANTES DE ENDOLISINAS FÁGICAS EN LIBRERÍAS METAGENÓMICAS MARINAS

Memoria de Trabajo Fin de Grado

Sant Joan d'Alacant

Diciembre 2016

**Autor:** Iris Fernández Ruiz

Modalidad: Revisión bibliográfica/experimental

Tutor/es: Dr. Francisco Rodríguez Valera

Dra. Ana Belén Martín Cuadrado

## ABSTRACT

In the recent years, the bacterial resistance against the current antibiotics has increased and it has become a recognized threat for the worldwide public health. Therefore, the search for new antibiotics and weapons to fight diseases caused by bacteria is a necessity that must be satisfied in the course of the following century. One option that is being studied by scientists nowadays are endolysins, proteins employed by bacteriophages to induce the lysis of the infected bacteria in order to release viral progeny. The laboratory of Francisco Rodríguez Valera (FRV) has performed the sequencing of a metagenomic library of fosmids from biomass of the prokaryotic community of the mediterranean DCM ("deep chlorophyll maximum") and new marine phage endolysins sequences were detected. At this work, 8 selected genes which codify putative endolysins were analyzed employing bioinformatic tools. Their domain-structures were studied and their sequences were compared with those from public databases. Holins and other lytic proteins were also screened in the fosmids where the endolysins were found.

En los últimos años la resistencia a los antibióticos actuales se ha incrementado drásticamente y se ha convertido en una amenaza para la salud pública mundial. Además, la búsqueda de nuevos antibióticos y arsenal terapéutico contra las infecciones bacterianas es una necesidad que se debe satisfacer a lo largo del próximo siglo. Una opción que está siendo estudiada actualmente por la comunidad científica son las endolisinas, proteínas empleadas por los bacteriófagos para inducir la lisis de la bacteria infectada y liberar a la progenie vírica. En el laboratorio de FRV secuenciaron una librería metagenómica de fósmidos de biomasa recogida en la DCM mediterránea ("*deep maximum chlorophyll*"), de donde detectaron varias secuencias de nuevas endolisinas fágicas marinas. En este trabajo se analizan 8 genes que codifican putativas endolisinas. Empleando herramientas bioinformáticas, se compararon con diferentes bases de datos públicas y se estudió sus dominios y posible estructura. Se realizó un análisis del entorno genómico de las endolisinas con el fin de buscar holinas y otras proteínas víricas líticas.

## Índice

1. INTRODUCCIÓN.....	1
1.1 Antecedentes	
1.2 Envoltura bacteriana y peptidoglicano	
1.3 Endolisinas	
1.3.1 Estructura de las endolisinas	
1.3.2 Tipos de endolisinas	
1.4 Otras enzimas líticas relacionadas	
1.4.1 Holinas	
1.4.2 “Pin-holinas”	
1.4.3 Spanins	
1.5 Mecanismo de acción de las endolisinas	
1.5.1 Bacterias Gram positivas	
1.5.1.1 Gram positivas Intracelulares	
1.5.2 Bacterias Gram negativas	
1.5.3 Micobacterias	
1.6 Aplicaciones en medicina	
1.6.1 Tratamiento de infecciones bacterianas	
1.6.2 Efectos adversos de endolisinas	
1.7 Otras aplicaciones	
1.8 Biodisponibilidad y farmacocinética	
1.8.1 Resistencias a las endolisinas	
1.8.2 Sinergia	
1.9 Enzibióticos y artilisinas	
1.10 Metodologías de estudio de endolisinas	
1.10.1 Estudio de actividad enzimática de endolisinas mediante Biología Molecular.	
1.10.2 Estudio <i>in vitro</i> de endolisinas, experimentación animal	
1.10.3 Estudio <i>in silico</i> de endolisinas, bioinformática y metagenómica	
1.10.3.1 Metagenómica – Librerías de fósmidos	
2. ANTECEDENTES PREVIOS A ESTE TRABAJO.....	15
3. OBJETIVOS.....	16
4. MATERIALES Y MÉTODOS.....	17
4.1 Obtención de las secuencias de endolisinas	
4.2 Características físico-químicas	
4.3 Comparación con secuencias conocidas	
4.4 Estudio estructural	
4.4.1 Péptidos señal	
4.5 Análisis del entorno genómico de las putativas endolisinas.	
4.5.1 Anotación de secuencias	
5. RESULTADOS.....	20
5.1 Resultados de la comparación genómica de los fósmidos completos	

5.2	Reanotación de secuencias	
5.3	Características físico-químicas	
5.4	Resultados de la comparación BLASTX de nuestras endolisinas con la base de datos “nr”.	
5.5	Estudio estructural	
5.5.1	Análisis de los dominios estructurales	
5.5.2	Predicción de estructura secundaria y terciaria. Estructura 3D-Resultados de Swissmodel	
5.5.3	Resultados de péptidos señal	
5.5.4	Resultados de la búsqueda de holinas y otras proteínas fágicas.	
5.5.5	Dominios transmembrana	
6.	DISCUSIÓN.....	30
7.	CONCLUSIÓN.....	34

**Palabras clave:** endolisina, muramidasa, antibacterianos, metagenómica, biología computacional, bacteriófagos, peptidoglicano.

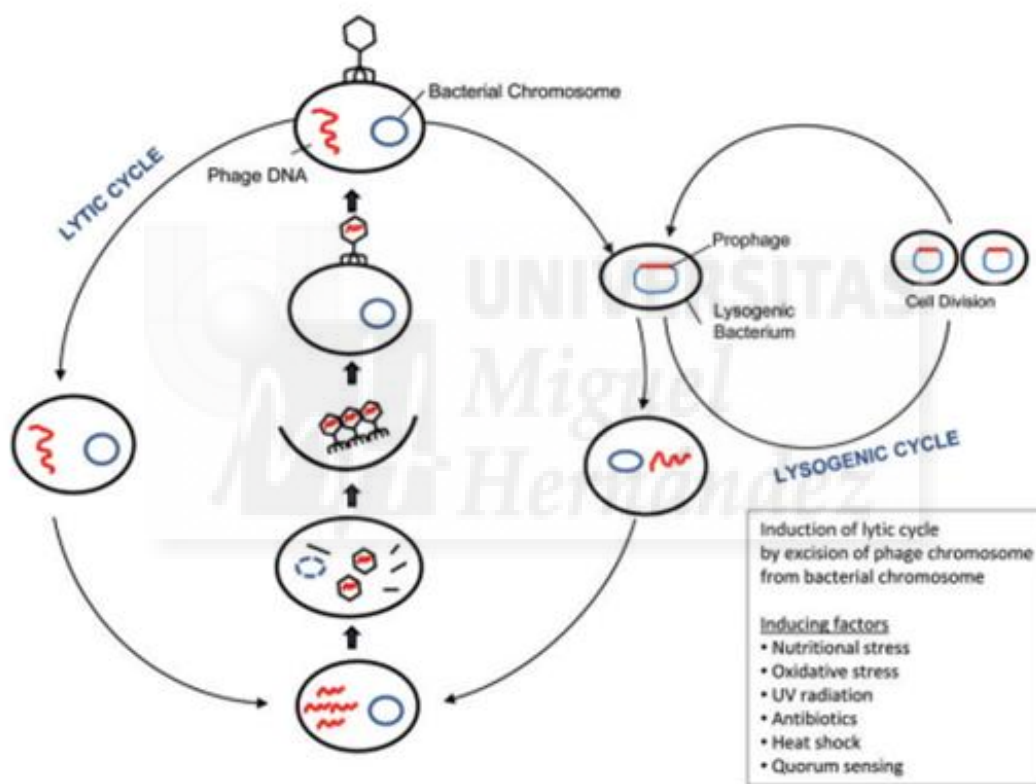
#### Abreviaturas utilizadas

- Peptidoglicano (PG)
- N-acetilglucosamina (GlcNAC),
- Ácido N-acetilmurámico (MurNAC)
- Membrana citoplasmática (CM)
- Peptidoglicano hidrolasas (PGH)
- Ácido desoxirribonucleico (ADN)
- Dominio de unión a la pared celular o “*Carbohydrate binding Domain*”, (CBD)
- Secuencia señal de secreción o “*signal-arrest-release*” (SAR)
- Deep Chlorophyll Maximum (DCM)
- Aminoácidos (aa)
- ORF (open reading frame)
- Secuencia codificante (CDS)
- No redundante (nr)
- Dominio peptidasa/aminohidrolasa cisteína, histidina-dependiente (CHAP)

# 1. INTRODUCCIÓN

## 1.1. Antecedentes

Se conocen como fagos a los virus que infectan específicamente a bacterias o archaeas. De manera general, los fagos pueden ser lisogénicos o líticos (Figura 1). Nuestro interés recae en la fase final del ciclo lítico cuando tiene lugar la lisis de la célula. Para ello, las proteínas fágicas, entre ellas las endolisinas, propician la ruptura de la célula huésped y la liberación de los viriones.



**Figura 1.** Ciclo lítico y lisogénico del bacteriófago<sup>(1)</sup>.

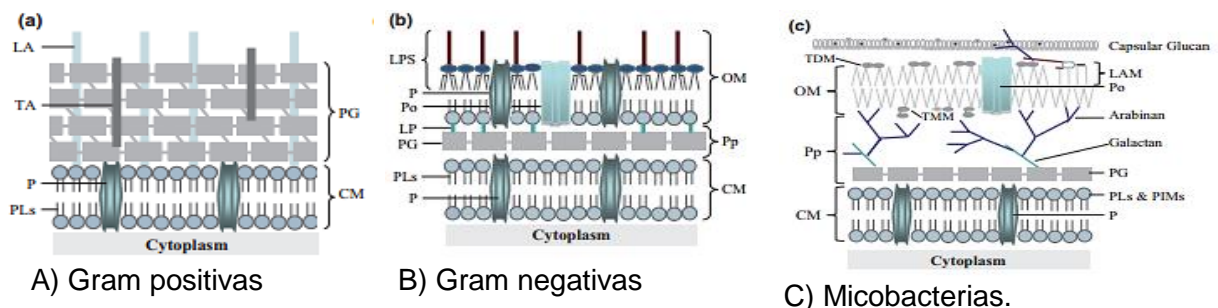
Las endolisinas fueron descubiertas en el siglo XX, pero su estudio y sus posibles aplicaciones es relativamente reciente debido al desarrollo de las técnicas bioinformáticas y el avance en el conocimiento de genomas de virus y bacterias. Actualmente, hay muchos estudios en desarrollo y descubrimiento de nuevas endolisinas debido a la escasez de nuevas dianas terapéuticas contra el desarrollo de infecciones causadas por bacterias patógenas resistentes a los antibióticos tradicionales.

## 1.2. Envoltura bacteriana y peptidoglicano.

Cada envoltura bacteriana es diferente, sobre todo en la composición de carbohidratos. Muestra de ello es, por ejemplo, la alta variabilidad de polisacáridos encontrados en *Salmonella* dando lugar a miles de serovares diferentes<sup>(2)</sup>.

De manera general, la pared celular bacteriana es la responsable de mantener la forma y la integridad de la bacteria frente a la presión osmótica. Uno de los componentes principales es el peptidoglicano (PG). El esqueleto del PG consiste en cadenas formadas por N-acetilglucosamina (GlcNAc) y ácido N-acetilmurámico (MurNAc) unidas por enlaces  $\beta$ -1,4 de GlcNAc. Las unidades de MurNAc tienen a su vez un enlace amida con cadenas peptídicas, las cuales forman puentes con otros péptidos unidos al MurNAc de otras cadenas de GlcNAc y MurNAc. La variedad de aminoácidos y entrecruzamientos es muy específica de cada bacteria.

Como vemos en la Figura 2, en Gram positivas la pared celular de PG es más gruesa que la de Gram negativas. Las bacterias Gram negativas poseen una doble membrana celular, con una fina capa de PG entre medias (Figura 2B). Las micobacterias tienen una envoltura compleja (Figura 2C), la membrana citoplasmática es parecida a la de otras bacterias, después hay un complejo formado por el PG de la pared celular, unido a arabinogalactano y a ácidos micólicos dándole un carácter hidrófobo; en el exterior tiene una capa formada por proteínas, polisacáridos y lípidos y por último una cápsula de glucano.



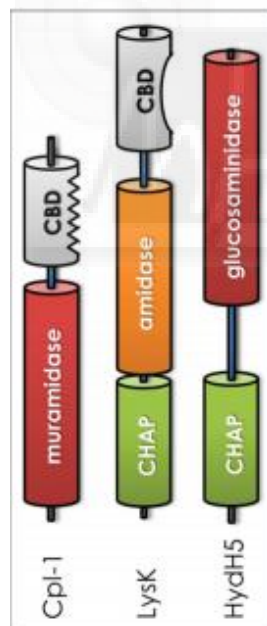
**Figura 2. Envoltura celular bacteriana: pared celular y membrana citoplásmica<sup>(3)</sup>.** CM: membrana citoplasmática, LA: ácido lipoteico, LAM: lipoarabinomanano, LP: lipoproteína, LPS: lipopolisacárido, OM: membrana externa, P: proteína, PG: peptidoglicano, PIMs: fosfatidilinositol manosido, PLS: fosfolípidos, Po: porina, Pp: periplasma, TA: ácidos teicoicos, TDM: trehalosa dimicolato y TMM: trehalosa monomicolato.

### 1.3. Endolisinas

Las endolisinas se pueden definir como enzimas víricas capaces de degradar el PG que compone la pared bacteriana. A grandes rasgos, se caracterizan por tener un dominio catalítico y un dominio de unión a la pared celular muy específico. Las endolisinas son peptidoglicano hidrolasas (PGH) y de manera general pueden diferenciarse dos tipos: (1) PGH asociadas a la cola del fago, como la chaperona endosialidasa, que actúan localmente en la inyección del ADN en el hospedador<sup>(4)</sup> y (2) endolisinas no asociadas a la cola del fago y que rompen el PG desde dentro de la célula en la fase tardía del ciclo de infección vírica, dando lugar a la liberación de los viriones formados.

Las endolisinas pueden ser purificadas y ser añadidas para lisar la pared de PG desde el exterior, lo que se conoce como “*lysis from without*” (a diferencia de la lisis desde el interior, “*lysis from within*”<sup>(5)</sup>).

#### 1.3.1 Estructura de las endolisinas



**Figura 3. Esquema de la estructura de endolisinas fágicas.** Dominio enzimático (muramidasa, CHAP) conectado a CBD, por una región flexible. Cpl-1 es una endolisina fágica que infecta a *Streptococcus pneumoniae*. LysK es una endolisina del fago K que infecta estafilococos. HydH5 es una PGH codificada por el fago phiPLA88 que infecta *Staphylococcus aureus*. (No representado a escala)<sup>(6)</sup>.

Aunque hay endolisinas globulares simples con un único dominio catalítico, éstas tienen generalmente estructura modular con dos dominios catalíticos (Figura 3). En el dominio N-terminal reside la actividad catalítica y en el C-terminal, la capacidad de unión a la pared celular (“*Carbohydrate Binding Domain*” o CBD). También es posible encontrar inversiones en este sentido,

endolisinas que no tienen dominio CBD y algunas que presentan hasta 3 dominios, siendo el tercer dominio también hidrolítico<sup>(7)</sup>.

### 1.3.2 Tipos de endolisinas

La clasificación de las endolisinas ha sido confusa durante varios años. Se les ha llamado tradicionalmente lisozimas y se clasificaban como tales en el mismo grupo que la lisozima humana y las autolisinas bacterianas. Hoy en día se les conoce como endolisinas o muramidasa porque rompen el PG o mureína. De forma general, se pueden clasificar en cuatro tipos generales según su actividad (Figura 4):

#### 1. Glicosidasa:

a. **N-acetilglucosaminidasa o glucosaminidasa:** rompe el componente glicano del PG, el enlace glicosídico  $\beta$ -1,4 que une GlcNAc y MurNAc, en el sitio reductor de GlcNAc. Es característico de autolisinas bacterianas pero se ha encontrado también en algunas endolisinas de bacteriófagos *LambdaSa2-like* como los que infectan estreptococos<sup>(8,9)</sup>.

b. **N-acetilmuramidasa o muramidasa:** rompe el enlace entre MurNAc y GlcNAc del PG, en el sitio reductor de MurNAc. Se refieren habitualmente como muramidasa o lisozima y se encuentra en autolisinas, exolisinas y endolisinas fágicas.

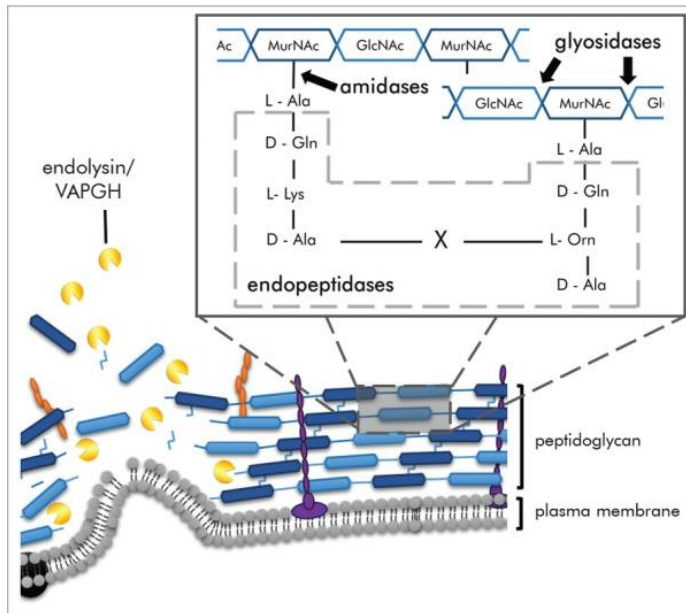
2. **N-acetilmuramyl-L-alanina amidasa o amidasa:** rompe el enlace peptídico entre MurNAc y el primer residuo de L-alanina del PG. Es típica de endolisinas fágicas.

3. **Endopeptidasa:** rompe enlaces entre dos de los aminoácidos que se encuentran unidos al MurNAc, hidrolizando el enlace amida.

4. **Transglicosilasa lítica:** no es una enzima hidrolítica como tal ya que no requiere la presencia de agua para romper el enlace PG. Rompen el enlace  $\beta$ -1,4 entre el MurNAc y el GlcNAc formando un residuo de N-acetil-1-6-anhidro-muramoil.

En la siguiente figura se puede ver la acción de las endolisinas sobre el PG.





**Figura 4.** Distintos tipos de actividad de las endolisinas sobre el PG <sup>(6)</sup>.

## 1.4. Otras enzimas líticas relacionadas

### 1.4.1 Holinas

Las holinas son proteínas hidrofóbicas pequeñas de menos de 150 aminoácidos, con al menos un dominio transmembrana (TMD) y con un dominio C-terminal hidrófilo<sup>(3)</sup>. Estructuralmente se clasifican en tipo I, II, III dependiendo de si tienen tres, dos o un TMD respectivamente. Se han clasificado un total de 7 superfamilias y su acción se ha visto que es necesaria en los fagos con ADN bicatenario<sup>(3,7)</sup>. El análisis de los genomas de fagos ha permitido averiguar que los genes que codifican para holina y endolisina suelen encontrarse adyacentes a las proteínas Rz y Rz1 formando el llamado “cassette de lisis” (Figura 5A)<sup>(7)</sup>. El “cassette de lisis” del fago Lambda ( $\lambda$ ) que infecta la bacteria *Escherichia coli* está compuesto por los genes S, R, Rz y Rz1. Estos codifican cinco proteínas: el R codifica la endolisina, los Rz1 y Rz codifican las “spanins” (ver más adelante) y, finalmente, el S tiene un “dual-start motif” que da lugar a dos proteínas opuestas, la antiholina S107 y la holina S105. Estas dos últimas se diferencian en que la antiholina tiene un extra de carga positiva, lo que previene la inserción del TMD en la membrana citoplasmática. Las antiholinas son degradadas por proteinasas y las holinas se liberan y acumulan en la membrana celular interna (Figura 4A).

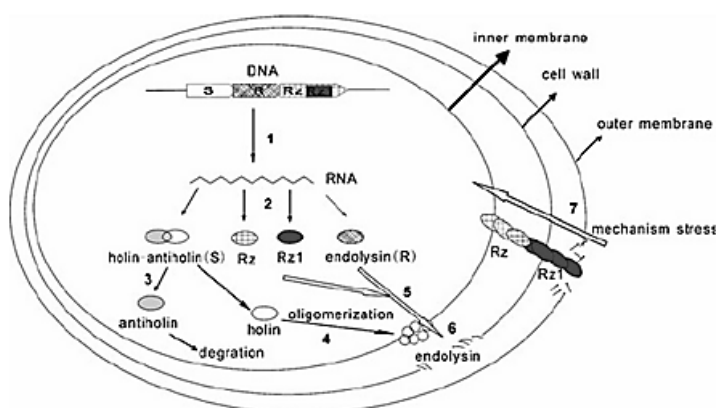
Las holinas se acumulan en el citoplasma hasta que llega el momento de realizar su acción, formando agujeros en la membrana citoplasmática y facilitando así la acción de las endolisinas en la pared bacteriana como vemos en la figura 5A y 5B. Por lo tanto, las holinas determinan el momento de la lisis celular, el cual es específico para cada tipo de fago y sólo ocurre cuando los viriones han terminado de formarse<sup>(3)</sup>. De hecho, se ha demostrado que si existen cambios de permeabilidad de la membrana provocados experimentalmente, su acción se realiza prematuramente<sup>(10)</sup>. En algunos fagos se necesita la presencia de una “*holin functional unit*”, un complejo de varios polipéptidos en vez de una sola proteína<sup>(3)</sup>.

#### 1.4.2 “Pin-holinas”

Las lisinas que no tienen secuencia señal para ser liberadas al exterior necesitan a las holinas<sup>(11)</sup>. Los fagos que producen endolisinas con secuencia señal de secreción, o “*Signal-arrest-release*” (SAR), sintetizan en paralelo proteínas chaperonas que interaccionan con ellas y son transportadas por la vía de translocación “*Sec*”<sup>(3)</sup>. En estos casos, los fagos codifican también para unas proteínas parecidas a las holinas, pero de menor tamaño, llamadas “*pinholins*”. Éstas, en vez de agujeros, forman canales heptaméricos pequeños que despolarizan la membrana (Figura 4B)<sup>(12)</sup>.

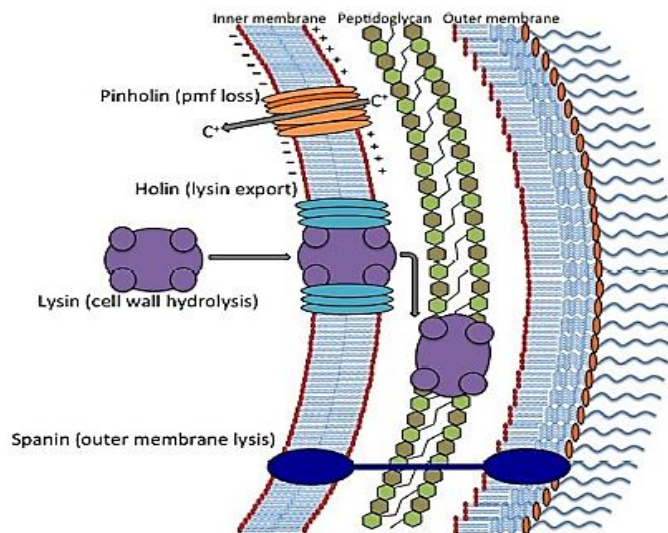
#### 1.4.3 “Spanins”

Son un complejo de proteínas lipoproteicas codificadas por los genes Rz1 y Rz (Figura 4A) debilitan la membrana exterior de las Gram negativas fusionando la membrana externa e interna (Figura 4B)<sup>(12)</sup>.



**Figura 5.**

**A)** “Cassette” de lisis del fago Lambda ( $\lambda$ ) y esquema de lisis celular originada por la acción de endolisinas<sup>(8)</sup>.



**B) Lisis en bacterias Gram negativas y las proteínas implicadas.** Las endolisinas producen hidrólisis celular, la holina ayuda a exportar hacia el exterior a la endolisina formando agujeros en la membrana celular interna, las pinholinas forman pequeños canales heptaméricos que producen pérdida de potencial de membrana y por último, en el caso de las Gram negativas, las “*spanins*” debilitan la membrana externa formando un Puente entre ambas membranas<sup>(12)</sup>.

## 1.5 Mecanismos de acción de las endolisinas

Dependiendo de si el fago tiene ADN de doble cadena o de cadena simple, la lisis se producirá de forma distinta:

1. **Los fagos de cadena simple** utilizan un único agente de lisis el cual actúa inhibiendo un paso de la biosíntesis del PG, por lo que se requiere que la bacteria esté en crecimiento<sup>(3,13)</sup>
2. **En los fagos de doble cadena** hay un sistema coordinado holina-endolisina, donde la holina forma poros en la membrana celular y la endolisina es la que rompe el PG de la pared celular. Como resultado, la célula es destruida por ósmosis.

Las endolisinas actúan sobre el PG, pero en función de si son Gram positivas o negativas, éste no está igual de accesible.

### 1.5.1 Bacterias Gram positivas

Las bacterias Gram positivas se caracterizan por tener una pared celular gruesa de PG (Figura 2A) y una rica composición en carbohidratos, los cuales son las dianas de las endolisinas. Se han descubierto diferentes tipos de endolisinas activas frente a Gram positivas, pero ninguna de ellas funciona con amplio espectro sobre todo el grupo, sino que se limitan a una especie bacteriana o incluso a una cepa concreta de bacterias ya que la composición química de la pared celular de cada una de ellas es muy diversa y específica de cepa. Por ejemplo, una endolisina puede ser activa frente a *Bacillus*

*anthracis* pero no serlo frente a otros *Bacillus*<sup>(9)</sup>. Esta es una de las principales razones por las que se requiere un mayor estudio genético de cepas patógenas; por lo que en el caso de utilización de endolisinas en la clínica no podrán ser utilizadas sin conocer previamente la bacteria causante de la infección.

#### 1.5.1.1 Gram positivas intracelulares

Se han descrito endolisinas, como PlyC (codificada por el fago C1), capaces de destruir la bacteria intracelular *Streptococcus pyogenes*, bacteria patógena causante de enfermedades graves y shock séptico. Al alojarse dentro de las células epiteliales humanas, en principio, debería de estar protegida frente a la acción de fagos. Sin embargo, se ha descubierto que la parte catalítica de esta endolisina está formada por dos subunidades y una de ellas es capaz de unirse a los denominados “*lipids rafts*” de las células epiteliales, que son zonas ricas en fosfolípidos y fosfatidilserina. PlyC entra en el interior celular por endocitosis y, una vez dentro, destruye a *S. pyogenes*.<sup>(14,15)</sup>

#### 1.5.2 Bacterias Gram negativas

Las bacterias Gram negativas tienen dos membranas celulares y entre ellas una fina capa de PG (Figura 2B). La lisis de las bacterias Gram negativas es un proceso muy regulado y se puede resumir en tres pasos: (1) Permeabilización de la membrana interna: en un momento específico, ésta se permeabiliza bien por una holina, o por una pinholina. (2) Degradación del PG por parte de la endolisina fágica. (3) Ruptura de la membrana externa gracias a la acción de las “*spaninas*”<sup>(10)</sup>.

#### 1.5.3 Micobacterias

La pared celular de las micobacterias es distinta al resto de bacterias y consiste en un complejo hidrofóbico de PG, arabinogalactano y ácidos micólicos (Figura 2C). Las micobacterias han coevolucionado con micobacteriófagos y, por lo tanto, éstos fagos son también capaces de penetrar en la célula y destruirla de manera muy similar a lo visto anteriormente para las bacterias Gram positivas y negativas. La mayoría de los fagos identificados hasta la fecha son de ADN de doble cadena<sup>(16)</sup>. Los micobacteriófagos producen la enzima LysA, que es una PGH, y LysB que rompe las uniones de

los ácidos micólicos a la capa de arabinogalactano. Se ha estudiado su uso sinérgico para combatir micobacterias patógenas como *Mycobacterium tuberculosis*<sup>(17)</sup>.

## **1.6 Aplicaciones en medicina**

El potencial biotecnológico de las endolisinas es muy alto y en la actualidad, tienen utilidad en diversos campos.

### **1.6.1 Tratamiento de infecciones bacterianas**

El problema actual de la resistencia a antibióticos es muy grave y ya se han encontrado resistencias contra antibióticos frente *S. aureus* de forma generalizada, resistencia a los carbapenemes para el tratamiento de *Klebsiella pneumoniae* y resistencia a fluoroquinolonas para el tratamiento de infecciones por *E. coli*. El tratamiento mediante quinolonas contra la gonorrea ha dejado de ser eficaz y, además, existe resistencia a la colistina, única alternativa para acabar con determinadas enterobacterias<sup>(18)</sup>. La Organización Mundial de la Salud ha elaborado un plan estratégico para intentar frenar las resistencias donde la utilización de fármacos que actúen con nuevos mecanismos de acción se hace indispensable. En la actualidad, se está estudiando el uso de endolisinas fágicas y endolisinas recombinantes, como la HY-133, en el tratamiento de infecciones de bacterias multi-resistentes como *S. aureus*, *Streptococcus* sp., *Clostridium* sp. y *B. anthracis*<sup>(19)</sup>.

### **1.6.2 Efectos adversos de endolisinas**

Se ha hipotetizado que el tratamiento de infecciones con endolisinas, al tratarse de proteínas, podrían desencadenar una respuesta inmune y ser neutralizadas, o podrían generar alergias graves. En los estudios de toxicidad de SAL-200, cuyo principio activo es la endolisina fágica SAL-1 del fago SAP-1 que infecta estafilococos<sup>(20)</sup> al administrarla más de una semana de forma intravenosa tanto en perros como en ratas se observó el inicio de la cascada de activación del complemento y la generación de anticuerpos contra la endolisina. Sin embargo, no se han observado anomalías en el sistema nervioso, el corazón o el sistema respiratorio<sup>(20)</sup>. En 2016 para el tratamiento de infecciones estafilocócicas se realizó un estudio en monos considerando tiempo menor de

una semana y fue bien tolerado a dosis única de 80 mg/kg por día o dosis múltiples de 40 mg/kg/día<sup>(21)</sup>.

### 1.7 Otras aplicaciones

**Biofilms.** Los biofilms son comunidades de microorganismos que producen una matriz extracelular y que están fuertemente adheridos a un tejido o superficie<sup>(5)</sup>. Son superficies de difícil acceso para la mayoría de antibióticos y, además, las bacterias que los forman suelen presentar resistencias a múltiples antibióticos y tener tiempo de duplicación altos. Este último aspecto es de gran importancia ya que las endolisinas tienen la ventaja de poder actuar sobre bacterias de crecimiento lento<sup>(9)</sup>. Las endolisinas como SAL-1, derivada del bacteriófago SAP-1 que infecta varios estafilococos, ha demostrado ser eficaz experimentalmente a este respecto. Otras están siendo analizadas para hidrolizar los componentes de la matriz extracelular o las bacterias que componen el biofilm<sup>(22)</sup>. Un buen ejemplo es la endolisina CF-301 (Contrafect®), la cual actúa sobre *S. aureus* y ha mostrado actividad frente a biofilms en combinación con antibióticos en animales con endocarditis. Sus ensayos de fase I finalizaron en el 2015 y su administración a voluntarios sanos fue bien tolerada y no tuvo efectos adversos. En la fase II del estudio se probará su eficacia en el tratamiento de bacteriemias resistentes<sup>(23)</sup>.

**Desinfectantes.** Las endolisinas se plantean también como desinfectantes en material quirúrgico debido a su gran actividad y la ventaja de no dejar residuos tóxicos y ser biodegradables.<sup>(9,24)</sup>

**Veterinaria.** En animales de granja, por ejemplo, se está estudiando el empleo de endolisinas intramamarias para tratar la mastitis bovina<sup>(25)</sup>. Para ello las endolisinas deben mantener su actividad lítica en presencia de leche<sup>(5)</sup>.

**Apicultura.** Como tratamiento de la enfermedad de “loque americano” producida por *Paenibacillus larvae* que afecta a las abejas *Apis mellífera*, se propone el uso de la endolisina PlyPI23 codificada por el fago phiIBB\_PI23 de *P. larvae* cuya actividad y seguridad ha sido demostrada<sup>(26)</sup>.

**Agricultura.** Uso de endolisinas en la prevención de enfermedades de plantas utilizando plantas transgénicas que expresen endolisinas, siendo así una alternativa a los bactericidas químicos<sup>(5,11)</sup>. Un ejemplo es la patata transgénica

que expresa la endolisina T4 efectiva frente a la bacteria *Erwinia carotovora*, causante de la “pierna negra”.

**Seguridad alimentaria.** En la industria láctica sería posible usar microorganismos fermentadores como *Lactococcus lactis* que secreten endolisinas mediante previa modificación biotecnológica<sup>(5)</sup>. Tal es el caso de *Lactobacillus casei* modificados que liberan endolisinas frente *S. aureus* durante el proceso de fermentación en la elaboración de queso<sup>(27)</sup>. Además, las endolisinas pueden inmovilizarse en diferentes superficies como nanopartículas fijadas en “films”, lo que abre un campo de investigación en materiales de embalaje para el transporte de productos alimenticios<sup>(5)</sup>.

Ya que las endolisinas presentan mayor afinidad que las interacciones antígeno-anticuerpo, pudiendo ser utilizadas en la detección rápida de bacterias de posibles contaminaciones en la cadena de producción alimentaria<sup>(28,29)</sup>. Se ha identificado actividad de diversas endolisinas y PGH quiméricas contra patógenos transmitidos por los alimentos como LysH5 para *Listeria* y biofilms de *Stafilococos* y *Lactococos*, Ctp1L para *Clostridium tyrobytyricum* o lambdaSA2-E-Lyso-SH3b para *S. aureus*, entre muchas otras. En este sentido, es importante que sean termo-resistentes, económicas y bioquímicamente compatibles con el alimento<sup>(5)</sup>.

**Descontaminación del medio ambiente:** en descontaminación de aguas contaminadas por cianotoxinas<sup>(30)</sup>.

**Aplicaciones en la producción industrial para la obtención** de compuestos semi-sintéticos a partir de compuestos producidos por bacterias, como ácidos grasos en la producción de biodiesel, reduciendo el coste del proceso y lo optimizaría respecto al uso de métodos tradicionales de ruptura celular<sup>(8)</sup>. En esta línea, las endolisinas también se podrían usar para lisar células en el laboratorio con el fin de obtener ácidos nucleicos y proteínas.

**Aplicación en la fabricación de vacunas.** Ya que las endolisinas rompen el PG, la bacteria muere por ósmosis dejando la membrana celular intacta. Estas estructuras pueden ser utilizadas para provocar la inmunización como alternativa a las vacunas atenuadas o inactivadas<sup>(24)</sup>.

## 1.8 Biodisponibilidad y farmacocinética

Las endolisinas presentan ventajas y desventajas en el tratamiento de infecciones frente a los antibióticos y la terapia fágica. La principal ventaja de las endolisinas, al igual que los fagos, es su especificidad de cepa bacteriana, por lo que actúan sobre la bacteria patógena deseada y la microbiota natural del huésped no resulta afectada. Un problema común de la terapia con fagos es que estos tienen una mala difusión por su tamaño respecto a las moléculas de antibióticos. Sin embargo, tiene la ventaja de que los fagos se replican en el lugar de la infección al replicarse en las células diana. Ya que el fago es específico de cepa, se usan “*cocktails*” de distintos fagos que actúan frente a varias cepas de la especie bacteriana.

La farmacocinética de las endolisinas es parecida a la de los antibióticos tradicionales<sup>(31)</sup>. Por ejemplo, en monos SAL200 la concentración máxima es dosis-dependiente<sup>(21)</sup>. Se han hecho estudios de inyección de endolisinas en animales y han resultado eficaces, aunque por el momento, principalmente se apuesta en la aplicación tópica de las mismas. Un buen ejemplo es Staphitekt™ (Microcos) para el tratamiento del acné en una línea de cremas y geles de uso tópico llamada Gladskin®.

Previo a su uso, las endolisinas deben ser aprobadas como fármacos biológicos al tratarse de proteínas<sup>(32)</sup>. Para la obtención de una molécula eficaz y efectiva se necesita una gran inversión por parte de la industria farmacéutica en estudios de fase I, fase II y fase III hasta su comercialización. Por ello, se está planteando en algunos países conceder licencias especiales<sup>(32)</sup>. Todavía no hay protocolos estandarizados internacionales para los estudios preclínicos como ocurre con los antibióticos (BSAC o EUCAST) para facilitar la reproducibilidad de los datos. Actualmente, la investigación se está llevando a cabo, principalmente, por parte de las universidades y pequeñas empresas.

### 1.8.1 Resistencias a las endolisinas

Los fagos en ciclo lisogénico pueden intercambiar fragmentos de ADN con el cromosoma bacteriano promoviendo resistencias o incluso virulencia, como ya se ha visto con cepas de *E. coli* enterohemorrágicas. En el uso de endolisinas, hasta ahora, no parece que se puedan generar resistencias cruzadas. Sin embargo, sí que es cierto que algunas cepas mutadas con



distinta pared bacteriana podrían evitar la acción de las endolisinas<sup>(9)</sup> si los sitios diana de las endolisinas se encuentran alterados<sup>(5)</sup>. Además, si la resistencia a las endolisinas ocurriera, sería posible diseñar nuevas PGH mediante ingeniería genética<sup>(6)</sup>.

### 1.8.2 Sinergia

Ya que las endolisinas matan bacterias independientemente de la fase de crecimiento en la que éstas se encuentren, éstas podrían ser compatibles con terapias de antibióticos bacteriostáticos. Éstas presentan sinergia cuando se utilizan con otras PGH<sup>(9)</sup>; por ejemplo, Cpl-1, endolisina del neumococo y gentamicina.

### 1.9 Enzibióticos y artilisinas

Se define un “enzibiótico” como un producto antibacteriano basado en la acción de una o varias enzimas. En el caso de las endolisinas, enzimas fágicas. Se denominan artilisinas a endolisinas modificadas por ingeniería genética con el objetivo de atravesar la membrana celular externa de las bacterias Gram negativas y llegar a destruir el PG de la pared celular de las mismas, consiguiendo así superar el obstáculo que impedía a las endolisinas completar su acción<sup>(17)</sup>. Art-175, una artilisina eficaz contra *Acinetobacter baumannii*, consiste en la fusión de una endolisina y una porción de un péptido desestabilizador de lipo-polisacáridos que facilita la translocación de la endolisina a través de la membrana<sup>(33)</sup>. Para aumentar su efectividad, se han estudiado los dominios de diferentes endolisinas y se diseñan artilisinas a partir de esa información y de los modelos informáticos de interacción endolisina-pared celular<sup>(34)</sup>.

Otro ejemplo de obtención de quimeras es invertir o combinar dominios de endolisinas o fusionarlas con otras proteínas<sup>(35)</sup>. Un ejemplo de esto es la endolisina B30 (que proviene de un fago que infecta estreptococos) y la lisostafina (producida por la bacteria *S. aureus* y utilizada como antimicrobiano contra estafilococos) para tener actividad frente a estreptococos y estafilococos<sup>(9)</sup>. También se han construido variantes de LysK (endolisina del fago *K* que infecta estafilococos) con lisostafina. También se ha propuesto fusionar péptidos “*cell-penetrating*” o dominios de transducción proteica que

tienen zonas cargadas positivamente con endolisinas para acceder a bacterias intracelulares.

## **1.10 Metodologías de estudio de endolisinas**

**1.10.1 Estudio de actividad enzimática de endolisinas mediante Biología Molecular.** Se utiliza PG purificado, el cual es hidrolizado por la endolisina a testar. Los fragmentos resultantes son analizados mediante distintas técnicas como el método Park-Johnson en el cual se mide la actividad glicosilasa<sup>(9)</sup>, o la secuenciación de productos de digestión de la actividad endopeptidasa o L-alanina amidasa; pudiendo también determinarse los sitios de escisión. En algunos casos, puede ocurrir que la endolisina presenta mayor actividad hidrolítica y espectro que el fago de donde provienen<sup>(9,36)</sup>. Otros métodos que existen son los ensayos de turbidez, los “zymogramas”, la observación de la lisis en placa y el estudio de la estructura cristalina<sup>(37) (9)</sup>.

### **1.10.2 Estudio *in vitro* de endolisinas, experimentación animal**

En la actualidad se están realizando estudios pre-clínicos en ratones con endolisinas que han demostrado su actividad enzimática y antimicrobiana. También existen ensayos en vacas para valorar la curación y prevención de la mastitis usando endolisinas. Staphitekt® para el tratamiento del acné ha sido aprobada en el 2016 para el tratamiento de infecciones de la piel en humanos, aunque todavía no puede ser utilizada en cirugía o heridas abiertas<sup>(38)</sup> (en la actualidad, en ensayos de fase I y II). Entre los microorganismos estudiados destacan *S. aureus*, *B. anthracis*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *S. pyogenes*, y de Gram negativas destaca *A. baumannii* y *Pseudomonas aeruginosa*. No se han descrito mecanismos de resistencia hasta la fecha y tampoco han aparecido mutantes tras dosis sub-letales *in vitro*<sup>(32)</sup>.

### **1.10.3. Estudio *in silico* de endolisinas, bioinformática y metagenómica**

La bioinformática y proteómica son necesarias para ofrecer nuevos dominios, quimeras, y diseño de endolisinas mediante ingeniería genómica<sup>(11)</sup>. Existen

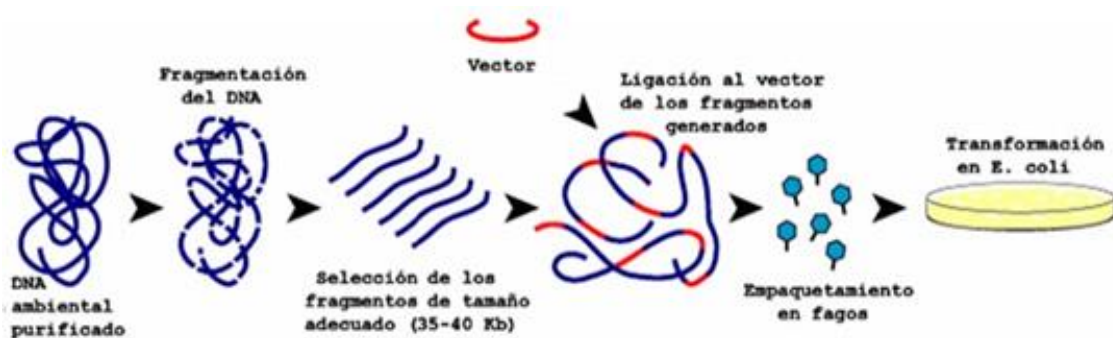
múltiples bases de datos genómicas que permiten comparar secuencias proteicas entre sí y predecir los dominios de las proteínas. La asignación del dominio catalítico se basa en la homología a otros dominios de endolisinas conocidas. A partir de la secuencia proteica y mediante estudios de mutagénesis dirigida se sabe qué residuos aminoacídicos están conservados y por lo tanto determinar los residuos activos del dominio catalítico<sup>(39)</sup>.

### 1.10.3.1 Metagenómica – Librerías de fósmidos

Se define metagenómica como el estudio del material genético recuperado directamente de muestras ambientales. Permite analizar el ecosistema en conjunto, lo cual adquiere una elevada importancia en los estudios ambientales. Mediante esta metodología es posible conocer la información genómica de microorganismos que son difíciles de cultivar (o incultivables, más del 99% en la naturaleza). Una variación del estudio del ADN ambiental es la creación de librerías metagenómicas construidas en vectores (fósmidos) que pueden contener hasta 35 Kb ADN. Estas librerías permiten buscar clones por secuencia o detectar la expresión de productos bioactivos como enzimas (“metagenómica funcional”)<sup>(40)</sup>. Este tamaño de fragmento de ADN es suficiente para poder contener un “cassette de lisis” fágico funcional<sup>(41)</sup>, por lo que las hace muy interesantes.

## 2. ANTECEDENTES PREVIOS A ESTE TRABAJO

El laboratorio de FRV secuenció en el 2012 más de 6000 clones de una librería metagenómica obtenida a partir de biomasa recogida en el Mar Mediterráneo a 60 m de profundidad (MedDCM-OCT2007)<sup>(42)</sup>. La metodología seguida fue la mostrada en la Figura 6. (1) Se recogió la biomasa de 60 litros de agua sobre filtros de 0.22 micras (usando prefiltros de 20 y 5 micras). (2) Se extrajo el ADN procariota y éste se clonó en fósmidos (pCCfos1, *CopyControl*<sup>TM</sup> *Fosmid Library Production Kit*, Epicentre). Se usó como hospedador *E. coli*. (3) Se secuenciaron los fósmidos. Entre las secuencias obtenidas, se obtuvieron decenas de nuevas secuencias de fagos que en el momento de obtención de la muestra estaban infectando a las células procariotas. Tras su ensamblaje y anotación, se detectaron varias ORFs que podrían codificar endolisinas fágicas. Ocho de ellas son las analizadas en este TFG.



**Figura 6.** Producción de bibliotecas metagenómicas mediante la utilización de fagos para transfectar el ADN clonado en *E. coli*.<sup>(41)</sup>

Actualmente, la búsqueda del término endolisina (“*endolysin*”) en la base de datos NCBI<sup>(43)</sup> da como resultado 20 resultados de dominios proteicos conservados, 52140 secuencias de proteínas, 311 de las cuales se han determinado experimentalmente su estructura molecular. Tan sólo han sido descritas dos endolisinas de origen marino<sup>(44,45)</sup>, la endolisina del fago GVE2, que infecta *Geobacillus* sp. E263 de las fumarolas hidrotermales de profundidad del océano del Este del Pacífico<sup>(44)</sup> y la endolisina del fago vB\_PmaS-R3 que infecta la cepa JL-65 de *Paracoccus marcusii* aislado de aguas superficiales del mar del Sur de China.<sup>(45)</sup>

### 3. OBJETIVOS

- Identificar la secuencia de putativas endolisinas presentes en la librería de fósmidos MedDCM-OCT2007 secuenciada en el laboratorio de FRV en el 2012.
- Realizar un análisis bioinformático de la estructura de las putativas endolisinas (dominios, estructura tridimensional y características físico-químicas).
- Conocer el entorno genómico de las secuencias anotadas como endolisinas. Búsqueda de holinas y otras proteínas fágicas asociadas a un putativo “*cassette* de lisis” que pueda existir en los fósmidos.

## 4. MATERIALES Y MÉTODOS

### 4.1 Obtención de las secuencias de endolisinas

Las secuencias de nucleótidos de los genes seleccionados para este trabajo se descargaron en formato FASTA desde <https://www.ncbi.nlm.nih.gov/>. La secuencia de aminoácidos también se descargó en paralelo (Tabla 1).

**Tabla 1.** Nomenclatura de las secuencias usadas en este trabajo.

Nomenclatura original	GenBank	Nombre en este trabajo
uvMED-CGR-U-MedDCM-OCT-S38-C3-cds30	BAQ94455.1	Proteína 1
uvMED-GF-C71-MedDCM-OCT-S28-C33-cds30	BAR30388.1	Proteína 2
uvMED-CGF-C117A-MedDCM-OCT-S27-C30-cds13	BAQ92505.1	Proteína 3
uvMED-GF-C104A-MedDCM-OCT-S27-C72-cds15	BAR25568.1	Proteína 4
uvMED-GF-U-MedDCM-OCT-S25-C63-cds19	BAR33194.1	Proteína 5
uvMED-GF-U-MedDCM-OCT-S32-C31-cds26	BAR33912.1	Proteína 6
uvMED-GF-U-MedDCM-OCT-S39-C22-cds37	BAR35268.1	Proteína 7
uvMED-CGR-U-MedDCM-OCT-S23-C23-cds22	BAQ89358.1	Proteína 8
uvMED-GF-U-MedDCM-OCT-S32-C31-cds51	BAR33937.1	Proteína 9

Estas secuencias, junto con otras muchas, fueron publicadas en el trabajo “*Expanding the marine virosphere using metagenomics*”<sup>(42)</sup> y su notación se realizó automáticamente. Un análisis posterior por parte de investigadores del grupo de FRV permitió la identificación de las primeras ocho secuencias de interés cuyo análisis informático se ha realizado en este TFG. La proteína 9 fue descubierta en este TFG. En la manipulación de las secuencias se utilizó el editor de texto PSPad<sup>(46)</sup>, ya que permite tratar los textos sin caracteres o formatos ocultos que pueden dar problemas posteriores en el análisis bioinformático. Así, se facilita el manejo de las secuencias en formato FASTA.

### 4.2 Características físico-químicas

Para predecir alguna de las características de las putativas endolisinas se usaron varios de los programas bioinformáticos accesibles desde ExPASy<sup>(47)</sup>, un portal de recursos bioinformáticos que da acceso a diferentes

bases de datos y herramientas. Para el análisis físico-químico se utilizó “ProtParam”<sup>(48)</sup>. Se calculó el índice alifático de la proteína (el volumen relativo de las cadenas alifáticas aumenta la termoestabilidad de proteínas globulares). Otro de los índices obtenidos fue el “GRAVY” o “promedio de hidrofobicidad”, que es la media de los valores de hidrofobicidad de los aminoácidos dividido por el número de aminoácidos de la secuencia. Se utilizó la herramienta “*Compute pI/Mw tool*” para conocer el punto isoeléctrico y la masa de la proteína estimada de la secuencia de aminoácidos <sup>(49)</sup>.

### 4.3. Comparación con secuencias conocidas

Se realizó una comparación de nuestras secuencias con la base de datos no redundante (“nr”, Genbank, NCBI) (formada por proteínas). Para esta comparación, se usó el programa BLASTX (parámetros predeterminados), el cual utiliza las secuencias de nucleótidos traducidas a aminoácidos para encontrar qué proteínas son similares a la secuencia que se consulta <sup>(50)</sup>. Ya que estas secuencias se encuentran ya publicadas, se descartó el primer resultado por ser ellas mismas y sólo se tuvieron en cuenta aquellos resultados con un  $E\_value < 10^{-5}$ . Este valor indica la probabilidad de que ese resultado sea por azar; por lo tanto, cuanto menor sea este, mejor. Además, se recopilaron datos de cobertura (“*query cover*”), el porcentaje de similitud y la descripción del producto de la proteína (la anotación y el organismo al que pertenece).

### 4.4 Estudio estructural

Para estudiar los dominios de las proteínas se utilizó el programa “Interpro” <sup>(51)</sup>, que usa bases de datos de dominio público del Instituto Europeo de Bioinformática. Este programa permite comparar una secuencia de aminoácidos con 11 bases de datos cruzadas entre sí con el objetivo de localizar dominios, sitios funcionales, realizar predicciones estructurales e incluso obtener información filogenética <sup>(52)</sup>. Las “*unintegrated signatures*” no están asociadas con entradas en Interpro, por tanto, no se tuvieron en cuenta. Para confirmar los resultados anteriores, se usó en paralelo el programa Pfam <sup>(53)</sup>. Para ello se usaron las secuencias en aminoácidos y sólo se analizaron aquellos resultados con un  $E\_value < 1.0$ .

Se usó el programa THMMM para calcular el número de dominios transmembranales (TMD)<sup>(54)</sup>.

Para el análisis de la estructura secundaria de las proteínas se utilizó el programa <http://bioinf.cs.ucl.ac.uk/psipred/> con el método PSIPRED v3.3. Ya que la predicción estructural tridimensional (3D) engloba los datos obtenidos de la estructura secundaria, se consideraron únicamente los obtenidos del análisis 3D en la discusión de los resultados. Para conocer la estructura en 3D, se utilizó Swiss-model<sup>(55,56)</sup>, el cual predice la posible estructura 3D de una proteína buscando similitudes en una extensa base de datos de proteínas cuya estructura se conoce. Ya que endolisinas víricas marinas no hay muchas descritas hasta la fecha, se aceptaron aquellas que tuvieran 20% de similitud.

#### **4.4.1 Péptidos señal**

Para la búsqueda de péptidos señal en las secuencia de las endolisinas putativas, indicativo de la secreción de estas proteínas al exterior celular, se utilizó el programa SignalIP<sup>(57)</sup>. Ya que desconocemos el posible hospedador del fago al que pertenecen estas proteínas, se usaron todas las opciones que presenta el programa usando como modelo tanto Gram positivas, como Gram negativas.

#### **4.5 Análisis del entorno genómico de las putativas endolisinas**

##### **4.5.1 Reanotación de secuencias**

Ya que la anotación de los fósmidos a los que pertenecen las endolisinas analizadas fueron anotados en el 2012, se procedió a re-anotarlos con una base de datos más moderna y con mayor cantidad de secuencias víricas conocidas. Se descargaron las secuencias de los fósmidos completos en formato FASTA del NCBI y se detectaron las posibles ORFs usando el programa Prodigal<sup>(58)</sup>. Las proteínas codificadas por los genes detectados en cada uno de los fósmidos se compararon mediante BLASTX. En este proceso, se usaron los servidores informáticos del grupo de FRV además de diversos programas desarrollados en este grupo de investigación.

Uno de los objetivos de este trabajo era la identificación de los “*cassettes* de lisis” a los que perteneciesen las endolisinas halladas. Para ello,

se estudió el entorno genómico de las secuencias anotadas como endolisinas, buscando también holinas y otras secuencias relacionadas. Como se ha explicado anteriormente, las holinas se caracterizan por tener varios TMD, por lo que se analizaron la presencia de todos los TMD de todas las proteínas del fósforo correspondiente. Para aquellas secuencias que tenían más de un TMD se analizó también un estudio de sus dominios en Pfam e Interpro para intentar obtener más información.

Se realizó una comparación genómica de todos los fósforos entre sí para ver su posible homología (aunque la representación de los genomas de los fagos es lineal, en realidad se trata genomas circulares). No se dispone de información sobre si el genoma del fago correspondiente se clonó completamente en el fósforo de la librería.

## 5. RESULTADOS

### 5.1 Resultados de la comparación genómica de los fósforos completos

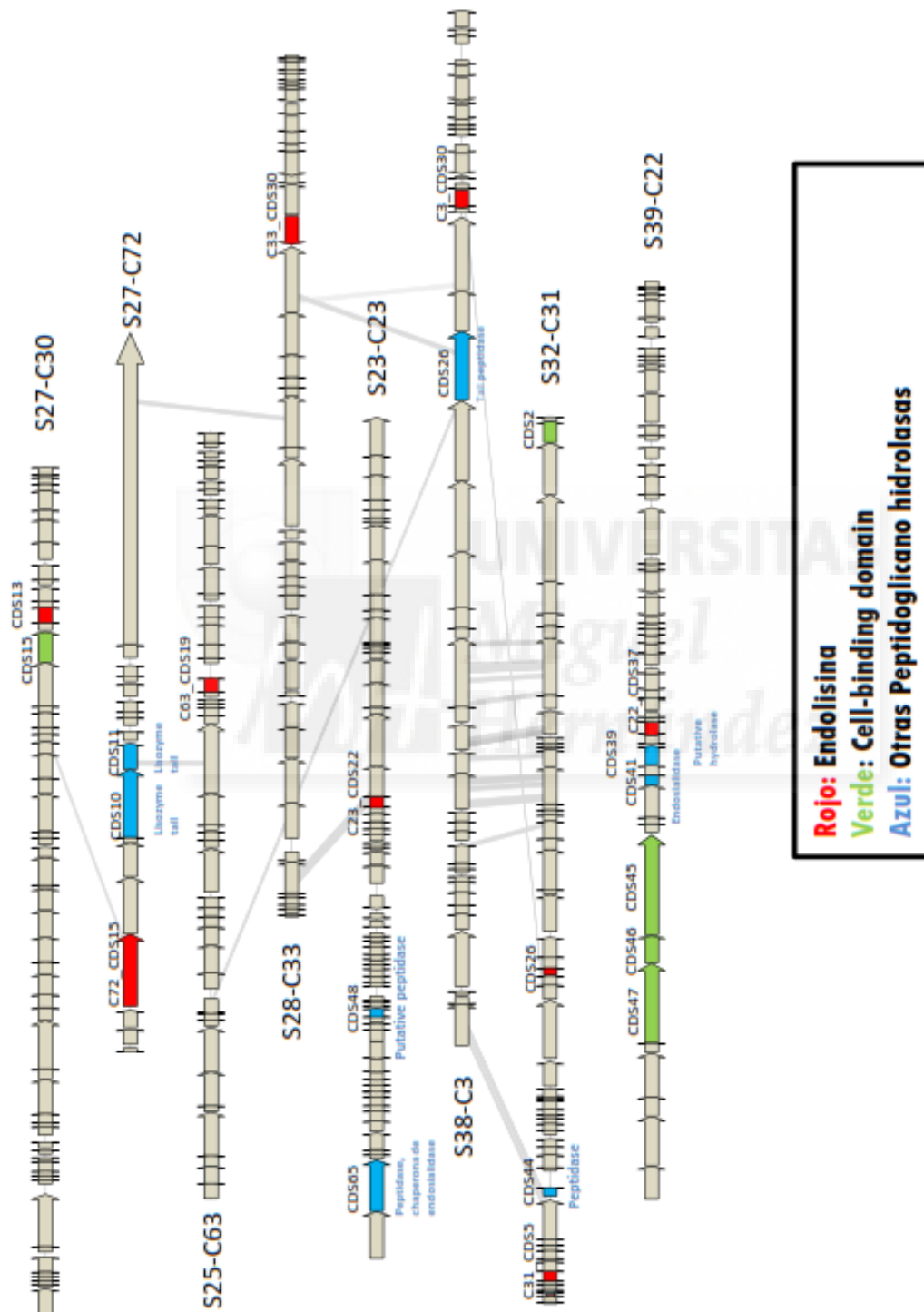
Los resultados indicaron que estos fagos son muy diferentes entre sí (Figura 7) y ninguno de ellos llegaban a compartir regiones de más de 1000 pb. Tampoco las endolisinas mostraron similitud entre ellas.

### 5.2 Reanotación de secuencias

En el fósforo S32-C31 se encontró una segunda posible endolisina codificada por la ORF S32-C31-cds51. Ésta fue analizada igual que las seleccionadas para este trabajo, llamándole en el resto de este trabajo “proteína 9”. Por lo tanto, la proteína 6 y la 9 pertenecen al mismo fósforo, pero al encontrarse distantes en el genoma, no parece que estén relacionadas. Las secuencias codificantes de los genes cds45, cds46 y cds47 del fósforo S39-C22 estaban anotadas como “proteína fibrosa, putativa lectina, hipotética proteína” respectivamente. Se vio en Pfam que estaban anotadas como “*Laminin\_G\_3\_Laminin\_G-2*” sugiriendo un putativo CBD. Se encontraron otras proteínas con capacidad hidrolítica, que podrían actuar localmente degradando la pared celular para inyectar el ADN del fago; así como otras PGH como peptidasas y endosialidasas <sup>(4)</sup>. En el fósforo S27-C30 se encontró una ORF anotada como CBD muy cercano a una secuencia codificante para una putativa



endolisina (proteína 3). En el fósforo S32-C31 se encontró una ORF, *cds2*, anotada como “proteína que contiene un dominio de unión a carbohidratos” y como “*CBM\_5\_12*” en Pfam, lo que sugiere que se trata de otro CBD, pero éste se encuentra alejado de la endolisina.



**Figura 7. Comparación genómica de los fósforos marinos que contienen las endolisinas putativas.** Cada flecha indica una CDS. Proteína 1 (S38-C3), 2 (S28-C33), 3 (S27-C30), 4 (S27-C72), 5 (S25-C63), 6 (S32-C31), 7 (S39-C22), 8 (S23-C23) y 9 (S32-C31). En azul, otras proteínas PGH fágicas no endolisinas (como *cds10* y *11* de S27-C72).

### 5.3 Características físico-químicas

Las putativas endolisinas que hemos analizado presentan un número de aminoácidos entre 98 y 401, a excepción de la Proteína 4, de 1019. El valor de punto isoeléctrico es diferente para cada una de ellas. Los valores negativos del promedio de hidrofobicidad indican ligera hidrofilia. El GRAVY presenta una relación con el índice alifático, cuanto mayor sea este último menos negativo será el primero.

**Tabla 2.** Características fisicoquímicas.

Proteína	Nº aminoácidos	Punto isoeléctrico	Peso molecular (Da)	Índice alifático	Promedio de hidrofobicidad (GRAVY)
Proteína 1	252	4,85	28369.79	71.67	-0.483
Proteína 2	401	5.65	44179.7	68.90	-0.525
Proteína 3	226	8.58	25222.87	71.59	-0.511
Proteína 4	1019	6.28	109893.7	75.87	-0.481
Proteína 5	205	6.97	23045.1	70.49	-0.317
Proteína 6	98	8.00	11022.7	114.39	-0.313
Proteína 7	187	7.77	20873.6	76.31	-0.367
Proteína 8	164	4.85	18120.7	94.63	-0.296
Proteína 9	143	4.80	16172.43	81.82	-0.284

### 5.4 Resultados de la comparación BLASTX de nuestras endolisinas con la base de datos "nr".

Se muestran en las siguientes tablas únicamente los dos resultados más relevantes obtenidos en la comparación BLASTX (15 de julio de 2016). En negrita se muestra la anotación de las proteínas a las que más se parecen nuestras endolisinas. Se muestra también el organismo al que pertenecen las proteínas que más se parecen y los datos de la cobertura, el E\_value y la identidad.

**Tabla 3. Resultados BLASTX proteína 1.**

Proteína 1	"Query cover" / E_value / Identidad	Anotación
Endopeptidasa LytF uncultured Mediterranean phage uvMED uvMED-GF-U-MedDCM-OCT-S42-C16 BAR36017.1	96% / 7e <sup>-65</sup> / 45% (242 aa)	Dominio tipo lisozima
Glúcosido hidrolasa, lisozima uncultured Mediterranean phage uvMED uvMED-CGR-U-MedDCM-OCT-S31-C1 BAQ94388.1	94% / 9e <sup>-48</sup> / 39% (244 aa)	Dominio tipo lisozima

**S38-C3-cds30**  
BAQ94455.1

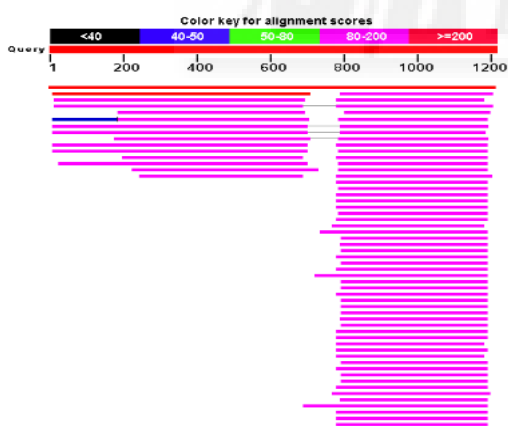
**Proteína 1.** Los resultados indican que la secuencia codificante para la endolisina es similar a dos secuencias anotadas como lisozimas de fagos

mediterráneos no cultivados (del mismo proyecto uvMED BioProject: PRJDB1484, <sup>(42)</sup> pero no estudiadas experimentalmente).

**Tabla 4. Resultados BLASTX proteína 2**

Proteína 2	"Query cover" / E_value / Identidad Anotación	S28-C33-cds30 BAR30388.1
lizozima putativa (Proteína lítica, muramidasa, endolisina) uncultured Mediterranean phage uvMED uvMED-GF-U-MedDCM-OCT-S39-C3 BAR35305.1	57%/ 2e-63 / 52% (229 aa) <b>Familia de tipo Peptidasa_C39</b>	
Lizozima "Synechococcus phage S-CBP2, dsDNA phage" YP_009103153.1	55%/ 1e <sup>-56</sup> / 48% (220 aa) <b>Familia de tipo Peptidasa_C39</b>	

**Proteína 2.** Los resultados de BLASTX indican que se parece a una endolisina de tipo muramidasa de fagos mediterráneos no cultivados (del mismo proyecto, pero no estudiadas experimentalmente <sup>(42)</sup> con una cobertura del 57% y una similitud del 52%. También, con una cobertura de 55% y una similitud del 48%, a una lizozima del fago S\_CBP2 que infecta *Synechococcus*. En el alineamiento se pueden observar dos zonas diferenciadas (ver Figura 8), es decir, hay proteínas que tienen mayor similitud con el extremo C-terminal de la proteína y otras que tienen mayor similitud con el dominio N-terminal; lo que indicaría la presencia de dos dominios diferentes.



**Figura 8.** Resultados de la comparación de la secuencia de la proteína 2 con las secuencias de proteínas de la base de datos "nr" mediante BLASTX. "Query sequence", en rojo, es la secuencia de consulta.

**Tabla 5. Resultados BLASTX proteína 3**

Proteína 3	"Query cover" / E_value / Identidad Anotación	S27-C30-cds13 BAQ92505.1
Proteína hipotética <i>Hassallia byssoidea</i> VB512170 Bacteria; Cyanobacteria; Nostocales; Microchaetaceae; <i>Hassallia</i> KIF31184.1	91% / 4e-55/ 44% (205 aa) <b>N-acetilmuramil-L-alanina amidasa</b>	
Hidrolasa de la pared celular "Photobacterium damselae" Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Photobacterium WP_065171825.1	87% / 5e-50/ 46% (208 aa) <b>Hidrolasa de la pared celular</b>	

**Proteína 3.** Los resultados de BLASTX muestran parecido con N-acetilmuramil-L-alanina amidasa y PGH de cianobacterias y de *Vibrios*.

**Tabla 6. Resultados proteína 4**

Proteína 4	"Query cover" /E_value/ Identidad Anotación	
Proteína hipotética uvMED-GF-C104A-MedDCM-OCT-S44-C56 BAR25574.1	99%/ 0.0/ 99% <b>NlpD DD-endopeptidasa de la mureína Peptidasa_M23</b>	<b>S27-C72-cds15</b> BAR25568.1
Proteína hipotética "Synechococcus phage S-CBP2 Viruses; unclassified dsDNA phages". YP_009103148.1	13%/ 8e-09/ 35% <b>Peptidasa_M23</b>	

**Proteína 4.** Los resultados del BLASTX muestran una identidad cercana, 99%, con una peptidasa de un fósido marino<sup>(42)</sup>. El segundo de los resultados mostró una similitud del 35% con una peptidasa del fago S\_CBP2 que infecta a *Synechococcus*.

**Tabla 7: Resultados BLASTX proteína 5**

Proteína 5	"Query cover" /E_value/ Identidad Anotación	
N-acetilmuramil-L-alanina amidasa, AmiC uncultured Mediterranean phage uvMED-CGR-U-MedDCM-OCT-S45-C18. BAQ94282.1	81% / 6e-30 / 40% <b>N-acetilmuramil-L-alanina amidasa, AmiC</b>	<b>S25-C63-cds19</b> BAR33194.1
Proteína hipotética <i>Vibrio parahaemolyticus</i> Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio. WP_025506093.1	88% / 6e-19 / 35%	

**Proteína 5.** Los resultados del BLASTX muestran similitud del 40% con una secuencia del fósido marino (S45\_C18) de actividad N-acetilmuramil-L-alanina amidasa. El segundo de los resultados mostró su parecido con una proteína hipotética de la *Vibrio parahaemolyticus* con una similitud de 35%.

**Tabla 8: Resultados BLASTX proteína 6**

Proteína 6	"Query cover" / E_value / Identidad Anotación	
Lisozima uncultured Mediterranean phage uvMED uvMED-CGR-C62A-MedDCM-OCT-S28-C10 BAQ94083.1	87% / 2e-14 / 43% <b>Proteínas con función de reconocimiento de PG (PGRPs)</b>	<b>S32-C31-cds26</b> BAR33912.1
Lisozima "Methylovorus sp. MM1 Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales; Methylophilaceae" Methylovorus.OAM52903.1	87% / 2e-12 / 36%	

**Proteína 6.** Los resultados del BLASTX muestran similitud con una secuencia del fósido marino S28\_C10, con proteínas con función de reconocimiento de

PG (PGRPs), del trabajo de Mizuno C. *et al.* <sup>(42)</sup>. El segundo de los resultados mostró su similitud con una lisozima bacteriana de *Methylovorus* sp. MM1.

**Tabla 9. Resultados BLASTX proteína 7**

Proteína 7	"Query cover" / E_value / Identidad Anotación	S39-C22-cds37 BAR35268.1
Lisozima "Synechococcus phage S-CBP3 dsDNA phage" YP_009103493.1	69% / 1e-50 / 65% <b>Familia de tipo Peptidasa_C39_2. Glucosaminidasa, Manosil-glicoproteína endo-β-N-acetilglucosaminidasa</b>	
Lisozima "Synechococcus phage S-CBP1 dsDNA viruses, Autographivirinae" YP_009103200.1	69% / 3e-50 / 65% <b>Tipo Peptidasa_C39. Glucosaminidasa; Manosil-glicoproteína endo-β-N-acetilglucosaminidasa</b>	

**Proteína 7.** Los resultados del BLASTX muestran similitud del 65% con lisozimas de actividad tipo glucosaminidasa, más concretamente Manosil-glicoproteína endo-β-N-acetilglucosaminidasa de los fagos SCBP3 y SCBP1 que infectan *Synechococcus*.

**Tabla 10. Resultados BLASTX proteína 8**

Proteína 8	"Query cover" / E_value / Identidad Anotación	S23-C23-cds22 BAQ89358.1
Proteína hipotética <i>Azospirillum brasilense</i> , Bacteria WP_014241523.1	69% / 3e-29 / 52% <b>MltE (transglicosilasa lítica soluble y proteínas relacionadas)</b> <b>LT_GEWL (Transglicosilasa lítica y lisozima "goose egg White")</b>	
Proteína hipotética <i>Polaromonas</i> sp. EUR3 1.2.1 Bacteria WP_026780159.1	67% / 2e-28 / 48%	

**Proteína 8:** Los resultados del BLASTX muestran similitud con transglicosilasas líticas solubles MltE y LT\_GEWL de *Azospirillum brasilense*, y una proteína hipotética bacteriana de *Polaromonas* sp.

**Tabla 11. Resultados BLASTX proteína 9**

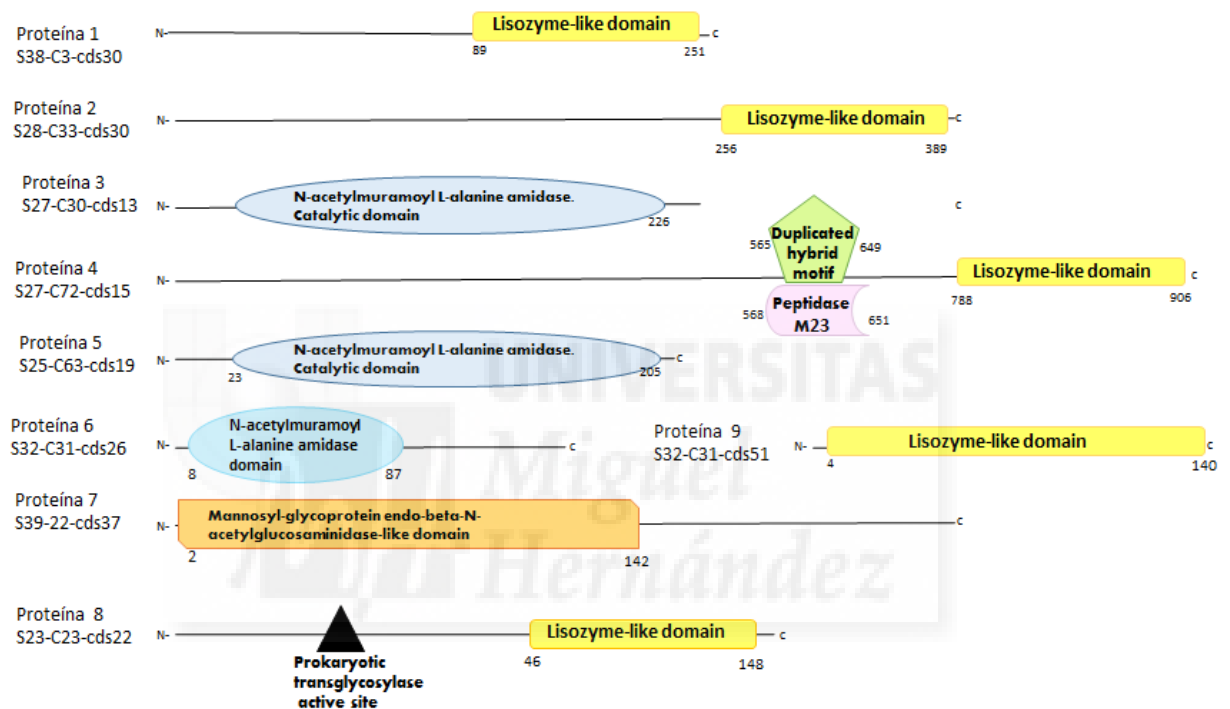
Proteína 9	"Query cover" / E_value / Identidad Anotación	S32-C31-cds51 BAR33937.1
Lisozima fágica (muramidasa) uvMED-GF-U-MedDCM-OCT-S26-C46 putative SAR11 Pelagibacter phage BAR32362.1	95% / 2e-56 / 66% (144 aa) Lisozima fágica ( <b>muramidasa</b> )	
Lisozima fágica (muramidasa) uvMED-CGR-C99-MedDCM-OCT-S44-C25 BAQ86098.1	95% / 1e-55 / 62% (141 aa) Lisozima fágica ( <b>muramidasa</b> )	

**Proteína 9:** Los resultados del BLASTX muestran similitud con muramidasa de un fago de *Pelagibacter* (SAR11) de muestra mediterránea S26-C46. También, con lisozima de muramidasa de otro fago del mismo muestreo, S44-C25, ambos del trabajo "*Expanding the marine virosphere using metagenomics*" <sup>(42)</sup>

## 5.5 Estudio estructural

### 5.5.1 Análisis de los dominios estructurales

Tal como se indicó en Materiales y métodos, se usaron dos programas diferentes, Pfam e Interpro. Los resultados obtenidos en ambos fueron similares. En la Figura 9 se muestran los dominios de las secuencias de las putativas endolisinas basada en la información obtenida de Interpro.



**Figura 9.** Dominios de las endolisinas putativas (no está representado a escala).

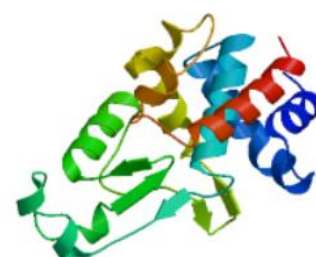
La proteína 1, 2, 4, 8 y 9 presentan un dominio tipo lisozima. Hay lisozimas de tipo glucosidasa y transglicosilasa. La proteína 2 y 9 pertenecen a la familia 24 de las glucosidasas, que incluye a la lisozima del bacteriófago  $\lambda$  que infecta a *E. coli*. En la proteína 8 hay un sitio activo para una transglicosilasa procarionte y, además, Interpro predice una función biológica de transglicosilasa lítica. En Pfam clasifica la proteína 9 como una lisozima fágica sin especificar el tipo de actividad lítica. Las proteínas 3, 5 y 6 son N-acetilmuramyl-L-alanina amidasas. La proteína 6 según Interpro y Pfam rompe específicamente el enlace amida de “*D-lactyl-L-ala*” y se encuentra en el

bacteriófago T7. En la proteína 4 encontramos una peptidasa M23, un “*duplicated hybrid motif*” además del dominio tipo lisozima, que indica que podría tratarse de una endolisina con dos dominios de actividad enzimática. Según Pfam, la peptidasa M23 es una Gly-Gly endopeptidasa y se une al Zn; dentro de la misma familia hay otras endopeptidasas y también lipoproteínas bacterianas. Según Interpro, el “*duplicated hybrid motif*” está presente en la PGH llamada LytM, un tipo de peptidasa M23. La proteína 7 tiene un dominio manosil-glicoproteína endo- $\beta$ -N-acetilglucosaminidasa que según Interpro y Pfam hidroliza el PG. Pertenece a la familia 73 de las glucósido hidrolasas. Es destacable que ninguna de las nuestras secuencias de putativas endolisinas parecen tener un dominio de unión a la pared celular bacteriana conocido.

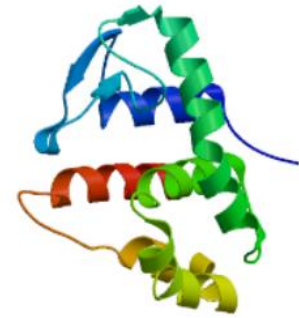
### 5.5.2 Predicción de estructura secundaria y terciaria. Estructura 3D-Resultados de Swissmodel.

Se muestran los resultados actualizados obtenidos el día 19/07/2016. En general, la similitud con proteínas cuya estructura se conoce fue muy baja, (siempre < 40%), probablemente por tratarse de proteínas de fagos marinos novedosas. Se tuvieron en cuenta similitudes, a veces remotas, por encima del 25% de similitud.

**Proteína 1.** La imagen a la izquierda corresponde al cristal del patrón 4ok7.1A, el cual tiene una identidad de 28,14%, una cobertura de 66% y corresponde a una endolisina fágica SPN1S. Ésta tiene un patrón inusual de dos módulos para la actividad lítica y de unión <sup>(59)</sup>. Es una endolisina rara que infecta a *Salmonella* y sobre la que se ha hecho estudios de unión a *E. coli* (Gram negativo).

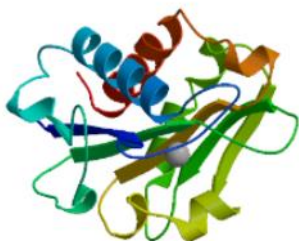


**Proteína 2.** El modelo 2anx.1.A explica el extremo C-terminal (cubre 1/3 de la secuencia proteica) y tiene una identidad del 37,14%. Corresponde a una lisozima mutante del bacteriófago P22. Tanto en la figura 3D como en la predicción de la estructura secundaria se observa un predominio de las  $\alpha$  hélices frente a las láminas  $\beta$  <sup>(60)</sup>.



**Proteína 3.** La proteína tiene un 23,81% de identidad con el modelo tridimensional 4m6g.1, el cual tiene actividad de tipo N-acetilmuramil-L-alanina amidasa y una cobertura de 84%. Es un monómero y tiene un ligando de zinc <sup>(61)</sup>.

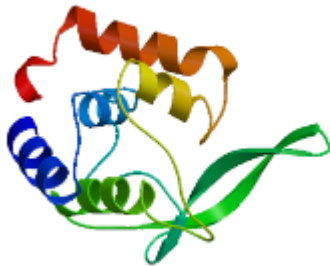
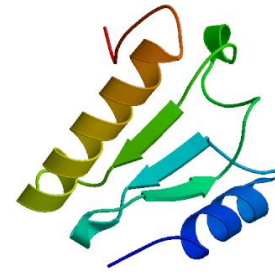
**Proteína 4.** La proteína 4 se parece a 4exm.1.B, con una cobertura de tan sólo un 16% y un 18,79% de identidad con la secuencia. Sin embargo, nos pareció interesante ya que 4exm.1.B es un cristal de una lisozima quimérica, donde el dominio de unión es de la pesticina (toxina de *Yersinia pestis*) y el catalítico corresponde a la lisozima del fago T4 . Es un homodímero <sup>(62)</sup> .



**Proteína 5.** La proteína 5 tiene una identidad del 25,15% con la secuencia de 1jwq.1.A, la cual es el dominio catalítico de la N-acetilmuramil-L-alanina amidasa de la bacteria *Paenibacillus polymyxa var.colistinus*, con una cobertura de 83%. Es un monómero <sup>(63)</sup> .

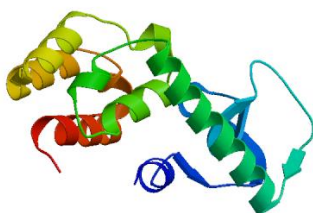
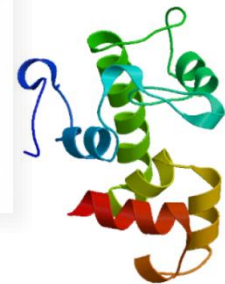


**Proteína 6.** La proteína presenta una identidad de 31,87% con el cristal 1lba.1.A, el cual es una lisozima del bacteriófago T7, con una cobertura de 93%. Es un monómero<sup>(64)</sup> .



**Proteína 7.** La proteína presenta una cobertura de 71% y una identidad de 30, 83% con la secuencia del patrón 3fi7.1.A, el cual corresponde al dominio catalítico de la autolisina de listeria. Es un monómero.<sup>(65)</sup>

**Proteína 8.** El modelo de 1qsa.1.A tiene una 29, 77% de identidad con nuestra secuencia en casi toda su longitud, cobertura (80%). Corresponde a la transglicosilasa lítica soluble de SLT70. Es un monómero.<sup>(66)</sup>



**Proteína 9.** El modelo de 191l.1.A tiene un 32,37 % identidad con la proteína 9 en un 97% de su longitud. 191l.A corresponde a una lisozima del bacteriófago T4. Es un monómero.<sup>(67)</sup>

### 5.5.3 Resultados de péptidos señal

Resultados de la búsqueda de péptidos señal usando el programa SignalP<sup>(57)</sup> usando como modelos tanto Gram positivas como Gram negativas son los siguientes (“*cut-off*” predeterminado):

**Tabla 12.** Resultados de péptido señal

Proteína	1	2	3	4	5	6	7	8	9
Gram negativas	No	Sí	Sí	No	No	Sí	No	No	No
Gram positivas	No	Sí	Sí	Sí	Sí	Sí	Sí	Sí	No

Todas tienen péptido señal excepto la proteína 1 y la 9.

#### **5.5.4 Resultados búsqueda de holinas y otras proteínas fágicas**

##### **5.5.1 Dominios transmembrana**

Se encontraron 26 ORFs con dominios transmembrana (TMD), con 1, 2, 3 y 4 TMD, de pequeño tamaño. Ninguno estaba anotado como holina, la mayoría eran proteínas hipotéticas. Se comparó su secuencia nucleotídica con de la base de datos de proteínas “nr” por BLASTX.

Del genoma del fósido S27-C30, de los ORF con dominios con TMD, cabe destacar el cds26 por presentar un péptido señal. Todas dieron lugar a proteínas anotadas como proteínas hipotéticas. Dentro del genoma del fósido que contiene la proteína 1, S38-C3, el cds36 que tenía un TMD era una supuesta fosfoesterasa llamada YqxK. Dentro del genoma del fósido que contiene la proteína 2, S28-C33, el cds43 (con dos TMD) codifica una proteína que contiene un dominio TM2 (Pfam) y un par de alfa hélices transmembrana unidas por un “*short linker*” (Interpro). Las proteínas codificadas por los cds42 y 43 del mismo fósido tienen un TMD y están anotadas como proteínas hipotéticas. Dentro del genoma del fósido que contiene la proteína 7, S39-C22, las proteínas codificadas por los cds35 y 36 tienen un TMD y están muy cerca de endolisina. Sin embargo, ambas son proteínas hipotéticas únicas y no se pudo extraer más datos sobre ellas.

## **6. DISCUSIÓN**

Las proteínas analizadas en este trabajo son putativas endolisinas de fagos del mar Mediterráneo. Hasta la fecha existen sólo dos endolisinas descritas de un hábitat marino, y ambas provienen de aislados de fagos de

bacterias marinas. Por tanto, existen pocos datos previos y resulta difícil su comparación con las ya existentes.

Se desconoce la composición de las envueltas celulares de muchas de las bacterias más abundantes en el océano y la posible variabilidad de los componentes de la pared celular intra-especie hace incluso más complejo este estudio. Esto explicaría por qué la cobertura y similitud con las secuencias disponibles en las bases de datos es tan baja (los parecidos con estructuras cristalinas eran incluso más bajos aún).

Los fósmidos en los que se encuentran estas endolisinas son muy diferentes entre sí, por lo que se deduce que pertenecen a fagos de diferentes grupos que infectarían a diferentes procariotas. Después de la re- anotación realizada de los mismos, no se pudo especular sobre la naturaleza de los posibles hospedadores. Un problema importante actual en muchas bases de datos es la anotación de secuencias. Por el volumen de datos genómicos, las anotaciones se hacen de manera automática basándose en las anotaciones previas. Por ello, es frecuente que un error de anotación se arrastre hasta la depuración/re- anotación de la misma. Además, es lógico pensar que a medida que avanza la ciencia, muchas proteínas hipotéticas dejen de serlo ya que, por ejemplo, se descubre su función. Por ejemplo, la cds22 del fósrido S23\_C23 fue anotada como una proteína hipotética inicialmente y en la reanotación realizada, sabemos que pertenece a la superfamilia de tipo lisozima. En otros casos, una misma secuencia se sube a la base de datos por duplicado (con distinto identificador de acceso) o bien cada científico le da un nombre diferente. Este ha sido uno de los mayores problemas en este TFG, ya que la misma endolisina puede estar anotada como lisozima, lisina o PGH. Por ejemplo, N-acetilmuramil-L-alanina amidasa y Ami\_3, como se ha explicado anteriormente en resultados, se nombra con diferentes sinónimos.

En la mayoría de las endolisinas, al analizar los dominios no indica la zona de CBD. Al re- anotar la secuencia de nuevo, se localizaron CBDs en alguna proteína de los genomas de fagos contenidos en los fósmidos analizados, pero tal como se ha comentado, estos genes estaban mucho de las endolisinas y muy probablemente, son proteínas implicadas en el reconocimiento de la diana del fago y degradación de la pared del huésped en el proceso de infección. Este es el caso de las cds45, el cds46 y el cds47 en el

genoma del fósido S39-C22 o las halladas en el fósido S32-C31. La excepción fue el fósido S27-C30, donde la endolisina (proteína 3) y el CBD se encuentran relativamente cerca, por lo que podría pensarse que hubo un error en la anotación y se trata de la misma ORF, o que son péptidos independientes pero que podrían traducirse conjuntamente y una hacer las funciones de CBD de la endolisina. Las proteínas con funciones de PGH que no fueron señaladas como endolisinas y se concluyó que podrían estar asociadas a la primera fase de infección.

Para la búsqueda de holinas se estudió los TMD presentes en todas las proteínas codificadas en el fósido. Se realizó una comparación de las secuencias con diferentes bases de datos, pero ninguna apareció anotada como holina, sino como proteínas hipotéticas o proteínas sin identificar. Por lo que sólo podemos especular que, aquellas holinas cerca de la endolisina, con TMD, puedan ser parte del “*cassette de lisis*” de la endolisina.

La predicción de los péptidos señal mostró que la mayoría de nuestras secuencias de endolisinas tienen un péptido señal. De ser así, se trataría de endolisinas tipo SAR, con secuencia señal de secreción que necesitarían chaperonas y serían transportadas por la vía de secreción celular.

No se pudo localizar en ninguno de los fósidos un “*cassette de lisis*” de las endolisinas típico.

Tras los análisis bioinformáticos, se concluye que:

**Proteína 1.** S38-C3-cds30. El análisis de dominios indica que es una lisozima, aunque no es posible especificar el tipo. La estructura 3D es similar a la endolisina fágica SPN1 que infecta a Gram negativas. En el fósido no se encontró CBD, por tanto parece una endolisina con un solo dominio catalítico, típico de endolisinas que infectan Gram negativas.

**Proteína 2.** S28\_C33\_cds30. El análisis de dominios indica que pertenece a la familia de las glucosidasas. En el fósido entero no se localizó ningún CBD, aunque al realizar el BLAST parecía que hubiera dos dominios diferentes. Se concluye que podría ser una endolisina con un solo dominio catalítico.

**Proteína 3.** S27\_C30\_cds13. Los resultados indican que tiene una actividad N-acetilmuramil-L-alanina amidasa. La proteína es similar a una peptidoglicano

amidasa de *M. tuberculosis*. Se localizó una secuencia codificante para CBD, S27\_C30\_cds15, separados por una ORF. La cercanía de la secuencia codificante para endolisina y la posible secuencia codificante para CBD plantean la separación de dichos dominios en dos genes diferentes, habiendo un ORF en medio. También es posible que no haya un codón stop y sea continuo. Se localizaron 7 TMD, de los cuales 3 eran adyacentes a la endolisina y tres CBD (cds11, cds12 y cds14) anotadas como proteínas hipotéticas, por lo que no se puede confirmar que sean holinas. Se concluye que también es una endolisina.

**Proteína 4** S27\_C72\_cds15. En el fósido entero no se localizó ningún CBD. Podría tratarse de una endolisina de tipo endopeptidasa con dominio catalítico.

**Proteína 5.** S25\_C63\_cds19. Tras los resultados del análisis de dominios y de la estructura en 3D se concluye que es una endolisina con un dominio N-acetilmuramil-L-alanina amidasa, el cual tiene un 25% de similitud con el dominio catalítico de N-acetilmuramil-L-alanina amidasa de *P. polymyxa var.colistinus*. No hay presencia de secuencias codificantes para CBD en el fósido, por lo que la endolisina puede tener un solo dominio catalítico.

**Proteína 6.** S32\_C31\_cds26. Por los resultados obtenidos se concluye que es una endolisina de tipo N-acetilmuramil-L-alanina amidasa (rompe “*D-lacyl-L-ala*”) típica del bacteriófago T7. El modelo de 3D confirma el parecido con la proteína de lisozima bacteriófago T7.

**Proteína 7.** S39\_C22\_cds37. Se concluye que la proteína 7 es una endolisina de tipo glucosaminidasa.

**Proteína 8.** S23\_C23\_cds22. Por los resultados del análisis de dominios y estructura proteica se concluye que podría tener actividad transglicosilasa lítica.

**Proteína 9.** S32\_C31\_cds51. Se concluye que es una glucosilasa del tipo del bacteriófago  $\lambda$ . El modelo de 3D mostró una similitud del 32% con la lisozima asociada a la cola del bacteriófago T4.

Fuera del ámbito este trabajo TFG, investigadores del grupo de FRV han continuado con el análisis de los fósidos donde se encuentran clonados estas endolisinas. El fósido al completo y/o la endolisina subclonada serán sometidas a diferentes pruebas para comprobar una su posible actividad.

## 7. CONCLUSIONES

En este trabajo se han analizado endolisinas anotadas automáticamente provenientes de una librería de fósidos de origen marino. Se ha realizado una predicción de su estructura en 3D, un análisis exhaustivo de sus dominios y se han predicho características fisicoquímicas de interés. Todas ellas presentaron características presentes en otras endolisinas descritas en literatura. También, se ha analizado el entorno genómico de las mismas y se ha podido encontrar CBD putativos y CBD asociados al reconocimiento de la pared bacteriana en la fase de infección. También se encontraron otras proteínas PGH fágicas. Los fagos representados por los genomas clonados en estos fósidos son todos diferentes entre sí. No se ha podido encontrar un “*cassette* de lisis” típico en ninguno de los casos. El avance en las anotaciones de nuevos genomas de fagos marinos permitirá mayor conocimiento de los mismos.



## REFERENCIAS

- 1- Knoll BM, Mylonakis E. Antibacterial bioagents based on principles of bacteriophage biology: an overview. *Clin Infect Dis* 2014;58(4):528-34. Epub 2013 Nov 23
- 2- Echeita MA, Aladueña AM, Díez R, Arroyo M, Cerdán F, Gutiérrez R, et al. Serotype and phage type distribution of human Salmonella strains isolated in Spain, 1997-2001. *Enferm Infecc Microbiol Clin* 2005;23(3):127-34 (In Spanish)
- 3- Catalao MJ, Gil F, Moniz-Pereira J, São-José C, Pimentel M. Diversity in bacterial lysis systems: bacteriophages show the way. *FEMS Microbiol Rev.* 2013; 37(4): 554-71. Epub 2012 Nov 8.
- 4- Hermoso JA, Garcia JL, García P. Taking aim on bacterial pathogens: from phage therapy to enzybiotics. *Curr Opin Microbiol.* 2007;10(5):461-72. Epub 2007 Sep 27.
- 5- Schmelcher M, Loessner MJ. Bacteriophage endolysins: applications for food safety. *Curr Opin Biotechnol.* 2016;37:76-87.
- 6- Roach DR, Donovan DM. Antimicrobial bacteriophage-derived proteins and therapeutic applications. *Bacteriophage*. [revista en internet] 2015 [fecha de consulta Julio 2016]; 5(3): p. e1062590. [aprox. 16 pantallas] doi: 10.1080/21597081.2015.1062590. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4590002/pdf/kbac-05-03-1062590.pdf>
- 7- Drulis-Kawa Z, Majkowska-Skrobek G, Maciejewska B, Delattre AS, Lavigne R. Learning from bacteriophages - advantages and limitations of phage and phage-encoded protein applications. *Curr Protein Pept Sci.* 2012; 13(8):699-722.
- 8- Gao Y, Feng X, Xian M, Wang Q, Zhao G. Inducible cell lysis systems in microbial production of bio-based chemicals. *Appl Microbiol Biotechnol.* 2013;97(16):7121-9.
- 9- Nelson DC, Schmelcher M, Rodriguez-Rubio L, Klumpp J, Pritchard DG, Dong S, Donovan DM. En: Łobocka M, Szybalski W. Chapter 7 Endolysins as antimicrobials. *Advances in Virus Research.* Editorial Elsevier, 2012;p299-365.
- 10- Young R. Phage lysis: do we have the hole story yet? *Curr Opin Microbiol.* 2013;16(6):790-7.
- 11- Fenton M, Ross P, McAuliffe O, O'Mahony J, Coffey A. Recombinant bacteriophage lysins as antibacterials. *Bioeng Bugs.* 2010;1(1):9-16.
- 12- Saier MH, Reddy BL. Holins in bacteria, eukaryotes, and archaea: multifunctional xenologues with potential biotechnological and biomedical applications. *J Bacteriol.* 2015;197(1):7-17.
- 13- Young R. Bacteriophage lysis: mechanism and regulation. *Microbiol Rev.* 1992;56(3): 430-481.
- 14- Gaca AO, Gilmore MS. A lysin to kill. *eLife*. [Revista en Internet]. 2016[31 agosto];5:[3 pantallas] <http://dx.doi.org/10.7554/eLife.16111> Disponible en <https://elifesciences.org/content/5/e16111>
- 15- Shen Y, Barros M, Vennemann T, Gallagher DT, Yin Y, Linden SB, et al. A bacteriophage endolysin that eliminates intracellular streptococci. [revisa en Internet] *eLife.* 2016 [31 agosto];5:[26 pantallas] Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4848087/pdf/elif-13152.pdf>
- 16- Pimentel M. Genetics of Phage Lysis. *Microbiol Spectr.* 2014;2(1): MGM2-0017-2013. doi:10.1128/microbiolspec.MGM2-0017-2013
- 17- Gerstmans H, Rodríguez-Rubio L, Lavigne R, Briens Y. From endolysins to Artilysin@s: novel enzyme-based approaches to kill drug-resistant bacteria. *Biochem Soc Trans.* 2016;44(1):123-8.
- 18- Organización Mundial de la Salud (OMS). Centro de Prensa. Resistencia a los antimicrobianos. Nota descriptiva. Organización Mundial de la salud, Ginebra. [Septiembre 2016;13 Octubre 2016]. Disponible en: <http://www.who.int/mediacentre/factsheets/fs194/es/>
- 19- Idelevich EA, Schaumburg F, Knaack D, Scherzinger AS, Mutter W, Peters G, et al. The Recombinant Bacteriophage Endolysin HY-133 Exhibits In Vitro Activity against Different African Clonal Lineages of the *Staphylococcus aureus* Complex, Including *Staphylococcus schweitzeri*. *Antimicrob Agents Chemother.* 2016;60(4):2551-3.
- 20- Jun SY, Jung GM, Yoon SJ, Choi YJ, Koh WS, Moon KS et al. Preclinical safety evaluation of intravenously administered SAL200 containing the recombinant phage endolysin SAL-1 as a pharmaceutical ingredient. *Antimicrob Agents Chemother.* 2014;58(4):2084-8.

- 21- Jun SY, Jung GM, Yoon SJ, Youm SY, Han HY, Lee JH, et al. Pharmacokinetics of the phage endolysin-based candidate drug SAL200 in monkeys and its appropriate intravenous dosing period. *Clin Exp Pharmacol Physiol*. 2016;43(10):1013-6.
- 22- Shen Y, Köller T, Kreikemeyer B, Nelson DC. Rapid degradation of *Streptococcus pyogenes* biofilms by PlyC, a bacteriophage-encoded endolysin. *J Antimicrob Chemother*. 2013;68(8):1818-24.
- 23- ContraFect Corporation. Nueva York, EEUU. 2016 [Acceso el 28 de octubre] Disponible en: <http://www.contrafect.com/technology/lysins>
- 24- Barrera Rivas CI, Cajero Juarez M, Oviedo-Boyso J, Nuñez-Anita RE, Kameyama Kawabe L, Valdéz Alarcón JJ. Advances in the Use of endolysins: general remarks, structure, applications, genetic modifications and perspectives. CI Barrera Rivas. Tarimbaro, Mexico:Editorial A. Méndez-Vilas; 2015 [17 de agosto 2016] Disponible en: <http://www.microbiology5.org/microbiology5/book/259-268.pdf>
- 25- Schmelcher M, Powell AM, Camp MJ, Pohl CS, Donovan DM. Synergistic streptococcal phage lambdaSA2 and B30 endolysins kill streptococci in cow milk and in a mouse model of mastitis. *Appl Microbiol Biotechnol*. 2015;99(20):8475-86.
- 26- Oliveira A, et al. The First *Paenibacillus* larvae Bacteriophage Endolysin (PlyPI23) with High Potential to Control American Foulbrood. *PLoS One*. [Revista en Internet];2015[15 julio 2016] 10(7): p. e0132095. [16 pantallas] <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0132095>
- 27- Guo T, Xin Y, Zhang C, Ouyang X, Kong J. The potential of the endolysin Lysdb from *Lactobacillus delbrueckii* phage for combating *Staphylococcus aureus* during cheese manufacture from raw milk. *Appl Microbiol Biotechnol*. 2016;100(8):3545-54.
- 28- Schmelcher M, Loessner MJ. Use of bacteriophage cell wall-binding proteins for rapid diagnostics of *Listeria*. *Methods Mol Biol*. 2014;1157:141-56.
- 29- Bai J, Kim YT, Ryu S, Lee JH. Biocontrol and Rapid Detection of Food-Borne Pathogens Using Bacteriophages and Endolysins. *Front Microbiol*. 2016;7:474. Epub 2016 doi: 10.3389/fmicb.2016.00474
- 30- Davison M, et al. Diversity in a Polymicrobial Community Revealed by Analysis of Viromes, Endolysins and CRISPR Spacers. *PLoS One*. [Revista en Internet] 2016[25 octubre];11(9): [16 pantallas] Disponible en:<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0160574>
- 31- Knoll BM, Mylonakis E. Antibacterial bioagents based on principles of bacteriophage biology: an overview. *Clin Infect Dis*. 2014;58(4):528-34. doi: 10.1093/cid/cit771. Epub 2013 Nov 23.
- 32- Cooper CJ, Khan Mirzaei M, Nilsson AS. Adapting Drug Approval Pathways for Bacteriophage-Based Therapeutics. *Front Microbiol*. [Revista en Internet] 2016 [5 Nov 2016];7:1209. [15 pantallas] doi:10.3389/fmicb.2016.01209 Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4971087/pdf/fmicb-07-01209.pdf>
- 33- Defraigne V, Schuermans J, Grymonprez B, Govers SK, Aertsen A, Fauvart M et al. Efficacy of Artilysin Art-175 against Resistant and Persistent *Acinetobacter baumannii*. *Antimicrob Agents Chemother*. 2016;60(6):3480-8. doi: 10.1128/AAC.00285-16
- 34- Briers Y, Lavigne R. Breaking barriers: expansion of the use of endolysins as novel antibacterials against Gram-negative bacteria. *Future Microbiol*. 2015; 10(3):377-90. doi: 10.2217/fmb.15.8.
- 35- Maervoet VE, Briers Y. Synthetic biology of modular proteins. *Bioengineered*. 2016:1-7. [Epub ahead of print]
- 36- Wang W, Li M, Lin H, Wang J, Mao X. The *Vibrio parahaemolyticus*-infecting bacteriophage qdvp001: genome sequence and endolysin with a modular structure. *Arch Virol*. 2016; 161(10):2645-52. doi: 10.1007/s00705-016-2957-x.
- 37- Keary R, Keary R, Sanz-Gaitero M, van Raaij MJ, O'Mahony J, Fenton M, McAuliffe, O et al. Characterization of a Bacteriophage-Derived Murein Peptidase for Elimination of Antibiotic-Resistant *Staphylococcus aureus*. *Curr Protein Pept Sci*. 2016;17(2):183-90.
- 38- Parmley S. Lysin in wait. *SciBX*. 2014;7(47)[2 pantallas] doi:10.1038/scibx.2014.1369 Disponible en: <http://www.nature.com/scibx/journal/v7/n47/full/scibx.2014.1369.html>
- 39- Linden SB, et al. Biochemical and biophysical characterization of PlyGRCS, a bacteriophage endolysin active against methicillin-resistant *Staphylococcus aureus*. *Appl Microbiol Biotechnol*. 2015;99(2):741-52. doi:10.1007/s00253-014-5930-1
- 40- Rodriguez-Valera F. Environmental genomics, the big picture? *FEMS Microbiol Lett*. 2004;231(2):153-8.



- 41- López López A, Zaballos Pérez M. Diversidad y actividad procariótica en ecosistemas marinos. *Ecosistemas*. 2005; 14(2): 30-40
- 42- Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using metagenomics. *PLoS Genet.*[Revista en Internet] 2013[citado 12 julio 2016].;9(12):13 pantallas Disponible en <http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003987>
- 43- National Center for Biotechnology Information (NCBI),[base de datos en Internet]. Bethesda:National Library of Medicine(USA) [sede Web]. Bethesda, USA; 1988 [Actualizado: 2016, Acceso:14 jul 2016] Disponible en:<http://www.ncbi.nlm.nih.gov>
- 44- Jin M, Ye T, Zhang X. Roles of bacteriophage GVE2 endolysin in host lysis at high temperatures. *Microbiology*. 2013;159(8):1597-605. Epub 2013 Jun 19.
- 45- Yongle X, Zhang R, Nianzhi J. Complete genome sequence of Paracoccus marcusii phage vB\_PmaS-R3 isolated from the South China Sea. *Stand Genomic Sci.* [Revista en Internet] 2015 [citado 14 nov 2016];10:94. [6 pantallas ] [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4641407/pdf/40793\\_2015\\_Article\\_89.pdf](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4641407/pdf/40793_2015_Article_89.pdf)
- 46- Fiala J. PSPad Freeware editor [sede Web] 2001[Actualización: 04 Jul 2016, acceso: 15 jul 2016] Disponible en: <http://www.pspad.com/es/download.php>
- 47- *Swiss Institute of Bioinformatics*, SIB. ExpASy Bioinformatics Resource Portal [Sede Web] Switzerland. 1993 Disponible en:<https://www.expasy.org/>
- 48- *Swiss Institute of Bioinformatics*, SIB. ExpASy Bioinformatics Resource Portal [Sede Web] Switzerland. 1993 [Acceso: 20 agosto 2016] ProtParam tool Disponible en: <http://web.expasy.org/tools/protparam/>
- 49- *Swiss Institute of Bioinformatics*, SIB. ExpASy Bioinformatics Resource Portal [Sede Web] Switzerland. 1993 [Acceso: 19 agosto 2016] Compute pI/Mw tool Disponible en: [http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)
- 50- Basic Local Alignment Search Tool (BLAST) [base de datos] Bethesda:National Library of Medicine (USA) National Center for Biotechnology Information, U.S. [Actualizado: 2 Jun 2016; Acceso:15 Jul 2016] Disponible en:<https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- 51- InterPro: protein sequence analysis & classification [Base de datos en Internet] European Bioinformatics Institute (EMBL) UK;1999, [Actualizada: 9 Junio 2016, Acceso: 15 Julio] Disponible en: <https://www.ebi.ac.uk/interpro/>
- 52- InterPro: protein sequence analysis & classification [Base de datos en Internet] European Bioinformatics Institute (EMBL) UK;1999, [Actualizada: 9 Junio 2016, Acceso: 15 Julio] Train online. Disponible en: <http://www.ebi.ac.uk/training/online/course/interpro-functional-and-structural-analysis-protei/summary>
- 53- European Bioinformatics Institute, Pfam [Base de datos en Internet] [Actualizada:1 July 2016, acceso 17 July 2016] Disponible en:<http://pfam.xfam.org/>
- 54- Center for biological sequence analysis. Technical University of Denmark. Dinamarca [Sede Web] [Actualizada: 4 Agosto 2015, Acceso: 20 Septiembre] Prediction of transmembrane helices in proteins TMHMM Server v. 2.0. Disponible en:<http://www.cbs.dtu.dk/services/TMHMM/>
- 55- Disponible en:<http://swissmodel.expasy.org/>
- 56- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*. [Revista en Internet] 2014 [citado 17 Jul 2016];42:W252-W258.Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4086089/>
- 57- Center for biological sequence analysis. Technical University of Denmark. Dinamarca [Sede Web] [Actualizada: 14 Marzo, acceso 11 Noviembre] SignalP 4.1 Server Disponible en: <http://www.cbs.dtu.dk/services/SignalP/>
- 58- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11(1):119. Disponible en:<http://prodigal.ornl.gov/>
- 59- Park Y, Lim JA, Kong M, Ryu S, Rhee S. Structure of bacteriophage SPN1S endolysin reveals an unusual two-module fold for the peptidoglycan lytic and binding activity. *Mol Microbiol*. 2014;92(2):316-325. Epub 2014 Mar 12 <http://onlinelibrary.wiley.com/doi/10.1111/mmi.12555/abstract;jsessionid=D7C855B71E1E6707CE8F5D75E7EC390A.f01t04>

- 60- Mooers BH, Matthews BW. Extension to 2268 atoms of direct methods in the ab initio determination of the unknown structure of bacteriophage P22 lysozyme. *Acta Crystallogr D Biol Crystallogr.* 2006;62(Pt 2):165-176.
- 61- Prigozhin DM, Mavrici D, Huizar JP, Vansell HJ, Alber T. Structural and Biochemical Analyses of Mycobacterium tuberculosis N-Acetylmuramyl-L-alanine Amidase Rv3717 Point to a Role in Peptidoglycan Fragment Recycling. *The Journal of Biological Chemistry.* 2013;288(44): 31549-31555.
- 62- Lukacik P, Barnard TJ, Keller PW, Chaturvedi KS, Seddiki N, Fairman JW et al. Structural engineering of a phage lysin that targets Gram-negative pathogens. *Proc Natl Acad Sci USA.* 2012; 109(25): 9857-9862.
- 63- Yamane T, Koyama Y, Nojiri Y, Hikage T, Akita M, Suzuki A, et al. The Structure of the catalytic domain of N-acetylmuramoyl-L-alanine amidase, a cell wall hydrolase from *Bacillus polymyxa* var. *colistinus* and its resemblance to the structure of carboxypeptidases. *Pendiente de publicación.*
- 64- Cheng X, Zhang X, Pflugrath JW, Studier FW. The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. *Proc Natl Acad Sci USA.* 1994;91(9):4034-4038.
- 65- Bublitz M, Polle L, Holland C, Heinz DW, Nimtz M, Schubert WD. Structural basis for autoinhibition and activation of Auto, a virulence-associated peptidoglycan hydrolase of *Listeria monocytogenes*. *Mol Microbiol.* 2009;71(6):1509-1522. Epub 2009 Jan 23. doi: 10.1111/j.1365-2958.2009.06619.x.
- 66- Van Asselt EJ, Thunnissen AM, Dijkstra BW. High resolution crystal structures of the *Escherichia coli* lytic transglycosylase Slt70 and its complex with a peptidoglycan fragment. *J Mol Biol.* 1999;291(4):877-898.
- 67- Zhang XJ, Baase WA, Matthews BW. A helix initiation signal in T4 lysozyme identified by polyalanine mutagenesis. *Biophys Chem.* 2002;101-102:43-56.

